Lonely cowboys? Lassoing happiness multiplier effects from unemployment and the measurement of group-level unemployment

Group 22

Abstract

We estimate a national and regional unemployment multiplier on life satisfaction of 7 and 4.5, respectively - the ratio of total societal welfare loss from unemployment to the contribution of those who are unemployed themselves. We show that using a familiar framework inspired by Di Tella and Oswald (2003) and an aggregated measure of unemployment using survey data, our estimated multiplier using country-level unemployment rates obtained from the IMF database is *twice* the size of the multiplier usually obtained in the literature. This finding is confirmed when we implement a new model specification technique using a double-selection procedure and LASSO, which identifies correct controls (micro and macro, squared and cubed terms as well as interactions) and thus minimises the risk of omitted variable bias and variance inflation. This new procedure identifies 13 relevant controls, only partly overlapping conventional choices. We also discuss how to test for the presence of no multiplier effect against a positive or negative effect.

This version: April 12, 2018

I. Introduction

In this paper, we estimate and examine the unemployment multiplier of life satisfaction. We estimate the multiplier on a national level to about 7, which is greater than what is usually found in the literature in the range of 3-5 (see e.g. Green (2011) and Di Tella and Oswald (2003)). We attribute this difference to two issues, which we study deeper in the paper.

The first is a measure of aggregate, group-level unemployment. This measure plays a crucial role in the estimation of the multiplier. In fact, we show that using a familiar framework inspired by Di Tella and Oswald (2003) and an aggregated measure of unemployment using survey data, the multiplier is only about *half the size* (and within the 3-5 range) of the multiplier obtained when we use country-level unemployment rates obtained from the IMF database. We also discuss how to test for the presence of no multiplier effect against a positive or negative effect.

The second issue is model selection. It is a strong pre-requisite for accurate measurement of the unemployment multiplier that *sufficient* and *adequate* controls are included in the model. By *sufficient*, we mean the inclusion of relevant control variables and its potential non-linear effects, such as interaction effects. Indeed, it is fair to presume that age and sex status would have polynomial and interaction effects with other control variables in the sample, respectively. By *appropriate*, we mean inclusion of only the relevant controls so that we avoid any unnecessary inflation of estimation error. Essentially, this issue can be compactly expressed as a matter of model uncertainty. The traditional approach in the literature has been to select a "favourite" model specification in which inference is conducted. This is typically accompanied by an sensitivity analysis in which the results for several different sets of controls are reported in an attempt to show that the effect of interest is insensitive to changes in the set of control variables. However, as shown in Belloni A. and Hansen (2014) and Belloni and Hansen (2017), the traditional model selection procedure will in many empirically relevant cases lead to invalid inference. This breakdown of traditional inference procedures essential occurs due to a failure to acknowledge the uncertainty involved in the model selection step. Acknowledging that the true specification of the model is unknown to the researcher, i.e. acknowledging that a oracle-type assumption of perfect model selection is unrealistic, the issue of model uncertainty and the avoidance of overfitting the model, we utilise a Least Absolute Shrinkage and Selection Operator (LASSO) type estimator as part of a double selection procedure as proposed recently Belloni and Hansen (2017). Henceforth, we refer to the procedure as LASSO Double Selection (LASSODS). In short, LASSODS is a two-step procedure that can be used to conduct robust inference about the effect of certain variables of interest. In the first step the feasible LASSO of Belloni and Hansen (2012) is utilized to choose which variables that should be included as controls, minimising the risk of omitted variable bias. In the second step the model is refitted using the variables selected in the previous step. The key feature of this method is that it does not rely on the oracle-type assumption of knowing exactly which control variables that should be included in the model. Instead, the feasible LASSO is used to select among a potential very large set of controls.[1] Under mild regularity conditions this procedure allows us to conduct valid inference about the unemployment multiplier. Furthermore, since the procedure (approximately) only includes the relevant control variables the standard errors will not be inflated. Additionally, the procedure is flexible, so that if the researcher will pre-determine some of the variables included as controls, the setup allows for this.

We generate large set of 68 variables, consisting of individual variables, squared and cubed terms as well as second-order interactions between both personal characteristics and macroeconomic factors. The LASSODS procedure identifies 13 variables deemed economically and statistically relevant, partly overlapping with conventional included variables. Using this set of control variables manifests that using the country-level, *external* measure of aggregate unemployment doubles the size of unemployment multiplier. It do, indeed confirm the findings above that conventional procedures produces a multiplier in the range of 3-5, whilst our new and robust double-selection LASSO procedures estimates an (oracle) multiplier of 7.

Lastly, we examine the effect own employment and family situation has on the

[1]The number of controls included in the first step can even be larger than the number of observations.

effect of other groups effect of life satisfaction. When estimated in our preferred models (the benchmark model and the double-selected model), we document an alleviating effect in the negative impact from others unemployment, when oneself is unemployed. We also document that marital status has no significant impact the effect of group-level unemployment. Finally, we compare the effect of regional-level and country-level groups effects of unemployment on individuals' life satisfaction. We find that the effect on a national level is generally greater than those on the regional level, but rarely on a statistically significant basis, suggesting only a modest further multiplication effect from regional to national level.

In Section II we operationalise the multiplier of unemployment on the basis of a social welfare functional approach and delineate its estimation procedure. Section III briefly discusses main attributes of the data and presents our measure of life satisfaction and unemployment metrics. In Section IV we concretise our implemented models and estimate the multiplier effect to answer the main question. We then answer the follow-up questions 1 and 2 separately in the empirical subsection C and D, respectively. The last follow-up question is answered in a separate discussion in Section V. We conclude in Section VI.

II. Operationalising the multiplier

In this section, we operationalise our understanding of the unemployment multiplier on life satisfaction and show how to estimate it.

A. A welfare functional approach

In order to clarify and define the multiplier effect of unemployment, we define the overall well-being in group society, g, inspired by Di Tella and Oswald (2003) as the utilitarian welfare

$$W_g = \left(1 - u_g\right) E_g + u_g V_g \tag{1}$$

where u_g is the unemployment rate within group g, E_g the utility of being employed in group g and V_g the utility of being unemployed in group g. Then by total differentiation, it follows that the total effect of a change in the unemployment rate in group g is

$$\frac{dW_g}{du_g} = \left(1-u_g\right)\frac{dE_g}{du_g} - E_g + u_g\frac{dV_g}{du_g} + V_g = \left(1-u_g\right)\underbrace{\frac{dE_g}{du_g}}_{(a)} + u_g\underbrace{\frac{dV_g}{du_g}}_{(b)} - \underbrace{(E_g - V_g)}_{(c)}$$

where (a) is the fear of unemployment for the employed while (b) is the fear of unemployment for the unemployed. The (c) is the personal cost of falling unemployed, i.e. the direct difference in utility of being employed or unemployed. Then a welfare functional multiplier effect is defined by the ratio between the total welfare loss, TE for "total effect", and the loss incurred just by the individual being unemployed, DE for "direct effect", such that $TE = IE + DE$, where IE is the residual "indirect effect", elaborated on below. The multiplier is given formally by

$$M = \frac{TE}{DE} = \frac{\frac{dW}{du}}{-(E-V)}. \tag{2}$$

Thus, if M>1 there may exist evidence of a 'fear effect', such that total effect on the overall group welfare is larger than the aggreation of the individual losses in welfare.

B. Estimating the multiplier

Suppose we have obtained a properly constructed measure of the i'th individuals' subjective well-being, which we define below, denoted by U_i. This is obtained from survey data and classified in X response categories. A natural choice of explanatory model for U_i is then the ordered probit model or the associated probit-adapted OLS model (POLS). A linear specification is postulated with the following pooled specification of the repeated cross-section

$$U_{ig} = \alpha + \beta u_{ig} + \gamma u_g + \delta x'_{ig} + \theta z'_g + \varepsilon_g + \lambda_t + \nu_{ig}, \tag{3}$$

where subscript g indicates a given group, for instance family, region or country. To fix ideas, let us supposes g denotes country. Then u_{ig} measures the individual level unemployment situation and u_g is the unemployment rate for the group as a whole. Given that we work with a repeated cross-section, we also include year dummies for two out of the three waives, denoted by λ. The control variables x_{ig} capture personal characteristics as income, age, education and the country-level control variables z_g capture other aspects of the macroeconomic stance such as inflation and GDP growth. If the model is estimated by POLS, estimated β and γ measure the marginal effects on individual level satisfaction by its personal unemployment situation and country-level unemployment rate, respectively. If the model is estimated by an ordered probit-model, marginal effects should be obtained via implied by the coefficient estimates.[2] Let $\hat{\beta}$ and $\hat{\gamma}$ measure the marginal effects obtained either from the ordered probit model or the POLS. On the basis of the theoretical groundwork put forward above, we may then define the estimated multiplier effect by

$$\hat{M} = \frac{\hat{\beta} + \hat{\gamma}}{\hat{\beta}} \tag{4}$$

To clarify, modeling the utility of being employed and unemployed in group g each as the conditional expectation, we can write the group welfare function as

$$W_g = u_g E\left(U_{ig}|x', u_{ig} = 1\right) + \left(1 - u_g\right) E\left(U_{ig}|x', u_{ig} = 0\right) = E\left(U_{ig}|x\right) \tag{5}$$

where $z = \left[x', u_{ig}\right]$ Our empirical model assumes that $E\left(U_{ig}|z\right)$ is linear in parameters. If we additionally assumes that the change in unemployment does not affect v_{ig}, our total effect of a change in unemployment becomes

$$\frac{dW_g}{du_{ig}} = \frac{dE\left(U_{ig}|z\right)}{du_{ig}} = \frac{\partial E\left(U_{ig}|z\right)}{\partial u_{ig}} \frac{du_g}{du_{ig}} + \frac{\partial E\left(U_{ig}|z\right)}{\partial u_{ig}} = \gamma \frac{du_g}{du_{ig}} + \beta \tag{6}$$

Let $du_{ig} = 1$,then the total effect in our empirical model becomes $\gamma + \beta$. Further, β is the partial effect of a change in unemployment cetirus paribus assuming that $du_g = 0$, i.e. ignoring spillover effects that could be capturing a 'fear effect'. To clarify, suppose $\hat{\beta} = -2/3$ and $\hat{\gamma} = -2$. Then the multiplier effect is $\hat{M} = 4$, indicating a fourfold multiplication effect.

[2] See e.g. Di Tella and Oswald (2003) for how to obtain marginal effects in a ordered probit model.

III. Data, measurement of happiness and unemployment

In this section, we briefly present our data and associated measures of well-being and unemployment.

We will focus on a sample of individuals in the age span of 21-60 who are part of the labour force at the time of the survey to be comparable to e.g. Clark, Knabe, and Ratzel (2010) who defined their sample in a similar way. Given our interest in the employment multiplier, this is a natural choice. We also sample macroeconomic data from all included countries in EVS except from Kosovo and Northern Cyprus (who constitute a limited share of the total sample) on real GDP growth[3]. Therefore the observations from Kosovo and Northern Cyprus is disregarded in the future analyses due. The collected macroeconomic data is obtained from the historical database of Maddison[4] and inflation rates from IMF database associated with the World Economic Outlook. Those measures were not included in the EVS, but are relevant for our model of interest to, among other things, control for macroeconomic effects such that the effect of the group-level unemployment measures is more precisely estimates. Furthermore, it would also give us a another measure of unemployment (in contrast to the measure also included in the analysis).

A. *Measure of well-being*

Our main measure of individual well-being is based on the following question

All things considered, how satisfied are you with your life these days?

Respondents could answer on a 1-10 point scale, where 10 corresponds to satisfied and 1 to dissatisfied. Empirically, several measures of subjective well-being are found to correlate well with each other, supported by factor analysis using both people's own judgement and other people's judgement of their overall well-being Frey (2008). Van Praag, Ferrer-i Carbonell et al. (2011) list the wild-spread use of this measure globally and highlight that most people in the same objective circumstance evaluate their life approximately by the same figure. This means

[3]We use the growth in GDP as opposed to levels to avoid potential issues with spurious regression arising from a unit root in the GDP data.

[4]See .

that our results will both be comparable to a range of other empirical studies, but also implies that this consistently captures a specific subjective measure, that at least to some degree can be translated across cultures.[5] Given cross-country (and cross-regional) nature of our data, we find these characteristics important.

Given the ordered nature of the data, we may estimate an ordered probit specification or a probit-adapted OLS (POLS). In the interest of simplicity, interpretability and since it is known to provide statistically indistinguishable marginal effects, we resort the to the computationally favourable POLS approach below.[6] To that end, we need to transform our life satisfaction data, which is bounded in [1,10], such that it is defined on $(-\infty,\infty)$, the entire real line. We simply achieve this by a normal transformation and center the data around zero, see e.g. Van Praag and Ferrer-i Carbonell (2004) and Van Praag et al. (2011) for details on the procedure.

B. Measure of unemployment

We define our group unemployment rate, u_g, as the share of unemployed in group g out of the total labour force in group g, where the labour force is defined to be the sum of the unemployed, the self-employed and the employed, both full-time and part time. This is the same definition used by OECD and the ILO. We will consider g as both a region and a country. In addition to the group unemployment rates based on the provided data, we sampled unemployment rates from all included countries in EVS (except Kosovo and Northern Cyprus) from the IMF database associated with the World Economic Outlook. Due to the omission of Kosovo and Northern Cyprus observations from these areas (who constitute a limited share of the total sample) will be discarded in the future analyses. This new unemployment measure also facilitates comparisons of the self-reported measure of unemployment that might contain measurement error.

[5] Some prefer measures such as suicide as a more valid measure, since it reflects revealed behaviour, but Bruno also argues that this only relates to tail behaviour.

[6] We do not take an explicit cardinalisation view of the our happiness measure, but note that application of the ordered probit or POLS implies an implicit cardinalisation by construction, see e.g. Van Praag and Ferrer-i Carbonell (2004).

C. *Missing data*

The EVS data set contains a non-negligible amount of missing values, both for the dependent variable, our well-being measure, and for the independent variables. We tested under the null hypothesis that the data was missing completely at random (MCAR) and rejected this hypothesis for almost all variables going into our analysis. Based on this we chose to use multiple imputation by chained equations, that allows us to handle continuous and discrete variables separately for imputing missing values for the independent variables, see e.g. Azur, Stuart, Frangakis, and Leaf (2011) for details. Missing values[7] are imputed by stochastic regression imputation type methods. E.g. binary variables are imputed by a logit model whereas ordinal data are imputed by ordered logit and so on. In brief terms, the procedure is an iterative procedure, that successively imputes missing values of a variable using other variables as predictors. Normally one would use an entire set of imputed datasets for a combined analysis, however, we chose to only advance with one imputed dataset [8]. We compared the distributions of the imputed values to the non-imputed values and found no discernible differences.

D. *Brief summary statistics*

We round this section of with some brief summary statistics. In Table 1 we report the distribution of responses to the well-being question posted above. Data is dispersed over all categories, but there is a greater mass associated with the upper half of the output interval. There is a clear monotonic and negative relationship between the degree of life satisfaction and unemployment, except form a peculiar effect in most satisfied group. Table 2 reports the country-wide distribution of well-being and unemployment. Generally Northern European countries are relatively more satisfied with their life and they experience, on average, less unemployment.

[7] Includes all the categories of 'Missing/Unknown', 'Not asked in the survey', 'Not applicable', 'No answer', and 'Don't know'

[8] In the interest of space and the time constraint, we refrain from this.

Table 1: Summary statistics of life satisfaction

This table reports summary statistics on the number of respondents and share of total respondents within each of the ten response categories of the life satisfaction measure. It also includes the unemployment rate (among the respondents), conditional on the response to the life satisfaction questions.

Life Satisf.	Frequency	Percent (%)	Self-reported unempl. rate (%)
1	1,984	2.78	34.17
2	1,258	1.76	26.87
3	2,917	4.08	22.49
4	3,373	4.72	18.03
5	7,795	10.91	15.32
6	6,848	9.59	11.48
7	11,627	16.28	8.84
8	17,374	24.32	6.56
9	10,079	14.11	5.51
10	8,177	11.45	7.64

Table 2: Summary statistics of life satisfaction

This table reports summary statistics on the number of respondents and share of total respondents within each of countries in our sample along with the average life satisfaction and the average unemployment rate over the three waves (which is self-reported) and the one obtained externally from the IMF World Economic Outlook database

Country	Frequency	Percent (%)	Life Satisf.	Self-reported unempl. rate (%)	IMF unempl. rate (%)
Albania	982	1.37	6.40	22.10	13.63
Armenia	709	0.99	5.66	30.32	18.03
Austria	2,239	3.13	7.74	4.06	4.05
Belarus	1,587	2.22	5.64	4.91	1.38
Belgium	3,226	4.52	7.50	12.00	7.37
Bulgaria	2,061	2.89	5.65	12.91	15.67
Canada	1,055	1.48	7.86	8.91	10.28
Croatia	1,396	1.95	7.07	17.84	17.18
Cyprus	485	0.68	7.36	5.98	5.10
Czech Republic	3,278	4.59	6.96	5.64	11.24
Denmark	2,178	3.05	8.39	5.69	6.18
Estonia	2,119	2.97	6.23	6.80	10.55
Finland	1,686	2.36	7.72	8.42	8.89
France	2,168	3.04	6.94	9.04	9.05
Georgia	869	1.22	5.67	53.74	16.57
Germany	4,220	5.91	6.95	11.04	7.14
Greece	1,372	1.92	6.89	7.29	10.74
Hungary	1,917	2.68	6.20	10.49	8.31
Iceland	1,671	2.34	8.09	2.33	3.82
Ireland	1,595	2.23	7.91	9.97	10.80
Italy	2,935	4.11	7.20	8.35	8.98
Latvia	2,050	2.87	5.86	10.98	10.66
Lithuania	2,106	2.95	5.90	10.54	12.76
Luxembourg	1,505	2.11	7.74	3.79	4.03
Macedonia	931	1.30	6.94	38.35	32.70
Malta	990	1.39	8.12	8.08	6.87
Moldova	858	1.20	6.66	30.65	5.93
Netherlands	1,866	2.61	7.95	2.89	4.17
Ireland	880	1.23	7.88	16.59	9.05
Norway	1,437	2.01	7.99	2.30	4.50
Poland	1,943	2.72	6.87	9.78	11.30
Portugal	1,797	2.52	7.04	10.02	6.33
Romania	1,861	2.61	6.12	9.30	6.23
Slovakia	2,343	3.28	6.73	10.12	18.24
Slovenia	1,971	2.76	7.09	9.54	8.30
Spain	2,631	3.68	7.26	11.48	16.71
Sweden	1,973	2.76	7.75	4.71	6.79
Switzerland	710	0.99	8.04	2.82	3.27
Turkey	1,349	1.89	5.59	29.50	9.76
Ukraine	1,450	2.03	5.55	15.66	9.49
United States	1,033	1.45	7.67	6.58	6.73

IV. Empirical results: The multiplier effect of unemployment

In this section, we delineate the specific model(s) we estimate and obtain estimates of the unemployment multiplier. We also provide empirical results on whether the effect of unemployment of others is larger, depending on the individual's own employment and family situation along with a comparison of regional and national effects.

A. Model specification

We consider at first the following country-level model, which is estimated by pooling the repeated cross-sections in the three waves of the EVS using POLS, and originates from the one briefly introduced in (3)

$$U_{ig} = \alpha + \beta u_{ig} + \gamma u_g + \delta x'_{ig} + \theta z'_g + \lambda_t + \nu_{ig}. \tag{7}$$

Given the nature of the data, we will adopt the simplistic view in which effects are permanent and there is not leads and lags effects. We estimate several specifications of the model in (7), which are reported in Table 3 below.

Table 3: Country-level probit-adapted OLS models

This table reports results from six different model specifications on a country-level framework; (1) and (2) do not include macroeconomic controls, but is elaborate on the personal characteristics. They differ only in terms of the measure of group-level unemployment rate, where (1) uses the aggreated self-reported measure and (2) used the external measure from IMP. Specifications (3) and (4), and (5) and (6) differ in a similar manner, whereas models (3)-(4) include macroeconomic controls, but is vague on the personal characteristic controls. The two last models includes both macroeconomic and personal characteristics as controls and is our preferred "benchmark model".

Cross Country Regression of Probit Adapted Life Satisfction						
	(1)	(2)	(3)	(4)	(5)	(6)
Unemp_i	-0.32***	-0.36***	-0.47***	-0.47***	-0.34***	-0.35***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Unemp_country(selfreported)	-1.40***		-0.91**		-1.00***	
	(0.48)		(0.34)		(0.34)	
Unemp_country(IMF)		-2.25**		-2.09***		-2.21***
		(0.97)		(0.74)		(0.76)
TE	-1.72	-2.61	-1.38	-2.56	-1.34	-2.56
DE	-0.32	-0.36	-0.47	-0.47	-0.34	-0.35
Multiplier Effect	**5.38**	**7.25**	**2.94**	**5.45**	**3.94**	**7.31**
Including Macroeconomic Variables						
Inflation			-2.12***	-2.22***	-2.11***	-2.22***
			(0.22)	(0.21)	(0.23)	(0.22)
Real GDP growth			-0.71	-1.22*	-0.81	-1.36**
			(0.63)	(0.63)	(0.61)	(0.60)
Including Personal Charecteristics						
Self-employed(1/0)	0.06*	0.04			0.06**	0.04*
	(0.03)	(0.04)			(0.03)	(0.03)
Male(1/0)	-0.03**	-0.03**	-0.02	-0.02	-0.03**	-0.03**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
More than 18 years of Education(1/0)	-0.01	-0.01			0.00	0.01
	(0.04)	(0.04)			(0.03)	(0.03)
Medium Income Level(1/0)	0.23***	0.23***			0.20***	0.21***
	(0.02)	(0.02)			(0.02)	(0.02)
High Income Level(1/0)	0.38***	0.38***			0.33***	0.33***
	(0.03)	(0.03)			(0.02)	(0.02)
Do you follow instructions at work?						
Can be convinced(1/0)	-0.10***	-0.10***			-0.08***	-0.07***
	(0.02)	(0.02)			(0.02)	(0.02)
Yes(1/0)	-0.11***	-0.10***			-0.09***	-0.09***
	(0.03)	(0.03)			(0.02)	(0.02)
Married(1/0)	0.13***	0.13***			0.18***	0.18***
	(0.02)	(0.02)			(0.02)	(0.02)
Age	-0.04***	-0.04***	-0.02***	-0.02***	-0.04***	-0.04***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Agesq	0.00***	0.00***	0.00***	0.00***	0.00***	0.00***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Constant	0.83***	0.86***	0.85***	0.94***	1.00***	1.09***
	(0.11)	(0.14)	(0.09)	(0.11)	(0.11)	(0.13)
N	74158	74158	74158	74158	74158	74158

We derive a number of interesting findings for the country-level analysis, but will for the sake of interest if space restrict ourselves to mostly commenting on the multiplier. A more elaborate analysis would investigate the estimates' implications for trade-offs and indifference curves. First of all, the multiplier is estimated convincingly above 1.We may test the hypothesis of no unemployment multiplication effect, i.e.

$$H_0 : M = 1, \tag{8}$$

simply by noting that $M = 1$ if and only if $\gamma = 0$.[9] That is, the null hypothesis in (8) may be rejected if estimated γ is significantly different from zero. We may test this by a conventional t-statistic using appropriate standard errors. It is clear from the Table 3 that we strongly reject the null of no multiplier effect of unemployment on life satisfaction against an increase in total welfare loss. These findings generally confirm what found in e.g. Green (2011) and Di Tella and Oswald (2003) of a non-trivial multiplication effect. However, we note a very interesting pattern. The potentially contaminated self-reported measure of group-level unemployment obtained from the EVS leads to a noticeably smaller multiplier than the use of a country-level measure obtained externally from the IMF database according to ILO standards. In fact, in our preferred benchmark models (5)-(6), the multiplier is alomost *twice* the size when using the external measure. We find this highly interesting and conjecture it may be associated with the presence of measurement error or, related, the fact that individuals' responses may be biased. As a conclusion, individual unemployment is related to high negative spillover effects.

We also consider the following regional-level model, where the aggregate unemployment rate is obtained from aggregation on the regional-level on the self-reported unemployment situation. We still include the country-level macroeconomic factors as controls. Table 4 reports the results.

[9] Technically, $M \to 1$ as $\beta \to \infty$ and $\gamma < \infty$. In practice, if the model is specified correctly, this is not relevant.

Table 4: Region-level probit-adapted OLS models

This table reports results from six different model specifications on a region-level framework. The specifications matches the numbering in Table 3.

Cross Region Regression of Probit Adapted Life satisfaction			
	(1)	(3)	(5)
Unemp_i	-0.37***	-0.48***	-0.37***
	(0.03)	(0.03)	(0.03)
Unemp_region	-1.40**	-1.28***	-1.19***
	(0.52)	(0.44)	(0.44)
TE	-1.77	-1.65	-1.56
DE	-0.37	-0.37	-0.37
Multiplier Effect	**4.78**	**4.46**	**4.46**
Including Macroeconomic Variables			
Unemp_country(selfreported)			
Inflation		-2.20***	-2.20***
		(0.27)	(0.12)
Real GDP growth		-1.07*	-1.23*
		(0.62)	(0.92)
Self-employed(1/0)	0.04		0.05
	(0.04)		(0.02)
Male(1/0)	-0.03**	-0.02	-0.03**
	(0.01)	(0.01)	(0.01)
More than 18 years of Education(1/0)	-0.00		0.02
	(0.04)		(0.02)
Medium Income Level(1/0)	0.22***		0.20***
	(0.02)		(0.02)
High Income Level(1/0)	0.37***		0.32***
	(0.03)		(0.02)
Do you follow the instructions at work?			
Can be convinced(1/0)	-0.11***		-0.08***
	(0.02)		(0.02)
Yes(1/0)	-0.10***		-0.09***
	(0.03)		(0.02)
Married(1/0)	0.11***		0.16***
	(0.02)		(0.02)
Age	-0.04***	-0.02***	-0.04***
	(0.00)	(0.00)	(0.00)
Agesq	0.00***	0.00***	0.00***
	(0.00)	(0.00)	(0.00)
Constant	0.80***	0.90***	1.01***
	(0.11)	(0.11)	(0.10)
N	38528	38528	38528

Standard errors in parentheses
Standard errors are clustered at region level
The regression is weighted according to population Weights
* p<0.10, ** p<0.05, *** p<0.01

The findings echo those of the country-level analysis, with a multiplier of roughly 4.5 (and statistically significant) and surprisingly constant across the three different model specifications. Unfortunately, we were unable to locate adequately granular external unemployment data to test whether the same results as in the national-level case holds true for the regional level. This is, indeed, an interesting area for future research.

B. An oracle multiplier: Lassoing the control variables via double-selection

From the former section, it stands out, not surprisingly, that the choice of control variables have a great influence on parameter estimates and, consequently, on the estimate of the multiplier. The reason is an omitted variable bias, where, among other things, the targeted variables (u_{ig} and u_g) may be correlated with elements in the error term. Our preferred model from above is generally elaborate and is consistent with the approach usually seen in happiness equation estimated in the literature (possibly without the macro component). However, inferences rely crucially on the assumption that we have specified the true model - an oracle like assumption. We suggest in this section a double-selection procedure, which not just robustifies inferences considerably, but also may alter conclusions drawn on the multiplier and parameter estimates in general. The following ideas are theoretically well-founded in a series of papers; Belloni and Hansen (2012); Belloni A. and Hansen (2014); Belloni and Hansen (2017). We refer to the Appendix below, where we detail some of technicalities and otherwise refer to the referenced papers. In this section, we will present the main ideas in an informal manner and the associated procedure.

The variable selection is in two-steps, using the Least Square Shrinkage and Selection Operator (LASSO). In a nutshell, LASSO estimates a linear model, for instance the one we consider in this article, by least squares but adds a penalty term. This penalty essentially puts a limit on budget of coefficients the estimator is allowed use, such that it conducts a selection procedure. LASSO possess the oracle property under mild assumptions, such that it asymptotically obtains the correct model with probability one. Denote by L_{ig} the set of candidate controls, both country-level, individual-level, squared and cubed terms as well as second-order interactions.[10] The double-selection then goes as follows:

1. Run a feasible LASSO and save relevant controls in L_{ig} for the dependent variable, U_{ig}.
2. Run a feasible LASSO and save relevant controls in L_{ig} for the target variables u_{ig}, u_g, individually.
3. Let the final set of control variables be given by the union of the controls selected in Step 1 and Step 2[11] and run the model defined above in (7) with those selected controls. Inferences can be conducted exactly as if the model was estimated by POLS in the beginning.

LASSO needs the choice of a tuning parameter which measures the importance of the penalty term. We choose this according to the recommendations put forward in Belloni and Hansen (2017), see also the Appendix. Since we will always include the year dummies and the target variable in our models, we always keep those in the models and run the feasible LASSO for selection purely among micro and macro controls and their second-order interactions. We have constructed a large set of 68 such candidate control variables and run the procedure just described. The resulting model along with the Step 3 parameter estimates are reported in Table 5 below.

[10] In principle, we could include higher-order interactions, but to maintain simplicity and clarity in interpretation, we assume second-order interactions are sufficient.

[11] If a variable is selected in bot step, we obviously include it only once.

Table 5: Double-selection LASSO probit-adapted OLS

This table reports results from the model specifications obtained by the double-selection LASSO procedure with the only difference between model (1) and (2) being the measure of group-level unemployment rate. The first model used the aggregated self-reported measure, whereas the second model uses the external unemployment rate from the IMF database

Cross Country Regression using LASSO control variables		
	(1)	(2)
Unem_i	-0.35***	-0.37***
	(0.03)	(0.03)
Unem_country(self reported)	-1.10***	
	(0.34)	
Unem_country(IMF)		-2.17***
		(0.73)
TE	1.45	2.53
DE	0.35	0.37
Multiplier Effect	**4.14**	**6.84**
Including macroeconomic variables		
Inflation	-2.04***	-2.10***
	(0.21)	(0.20)
Including personal characteristics		
Self-employed	-0.01	-0.01
	(0.03)	(0.03)
Male(1/0)	-0.05**	-0.05**
	(0.02)	(0.02)
Income	0.14***	0.14***
	(0.01)	(0.01)
Do you follow the instructions at work?	-0.08***	-0.07***
	(0.02)	(0.02)
Yes(1/0)	-0.09***	-0.09***
	(0.03)	(0.03)
Married	0.13***	0.14***
	(0.02)	(0.02)
Age	-0.01***	-0.01***
	(0.00)	(0.00)
Head of Household(1/0)	-0.07*	-0.08**
	(0.04)	(0.04)
Interactions		
Married*Head of Household	0.03	0.03
	(0.02)	(0.02)
sex*Head of Household	0.08**	0.08**
	(0.03)	(0.03)
Selfemployed*Head of Household	0.03	0.03
	(0.04)	(0.04)
Do you follow instructions at work?		
Can be convinced(1/0)*Head of Household	0.00	0.00
	(0.02)	(0.02)
"Yes(1/0)*Head of Household"	-0.02	-0.02
	(0.02)	(0.02)
Inflation*Selfemployed	0.32	0.07
	(0.21)	(0.17)
Constant	0.39***	0.49***
	(0.08)	(0.10)
Include occupation dummies	YES	YES
Observations	74158	74158

Standard errors in parentheses
Standard errors are clustered at country level
The regression is weighted according to population weights
* p<0.10, ** p<0.05,*** p<0.01

The resulting set of controls consists of 13 variables. Some are personal characteristics and some are interactions with inflation as well as inflation itself. Generally, the double-selection procedure seems to be attracted to interaction terms, but are not as popular in the literature. It is interesting that the estimated multiplier is, again, notably higher when using the external, IMF database for measuring the unemployment rates as compared to the aggregated self-reported measure from EVS. In a nutshell, the double-selection procedure robustifies the notion that the multiplier is at a level of about 4 on the national level when using a self-reported measure for aggregate unemployment. It also supports the finding from above that using a different, external measure of the aggregate level of unemployment leads to a multiplication effect of about 7.

Summarising above findings, we convincingly estimated a national multiplier of unemployment at about 7, which differs interestingly from the literature. This is obtained in both a standard framework and in a robustified framework using the double-selective LASSO procedure.

C. Follow-up question 1: Are the effects of the unemployment of others larger depending on own employment and own family situation?

In this section, we examine whether the effects of the unemployment of others are larger, depending on own employment and own family situation. We will do this by means of our preferred "benchmark" model and the one using the LASSO double-selection,[12] both in the regional and national level. To that end, we define the two interaction components

$$q_{1ig} = u_g \cdot u_{ig} \quad \text{and} \quad q_{2ig} = u_g \cdot family_{ig}, \tag{9}$$

where $family_{ig}$ measures the individual's own family situation. We examine the following family-related measures

- $familiy_1$: This survey question asks about 'is it important that children work hard?' on a binary scale. This is used to proxy parental expectations

[12] Unfortunately, we did not manage to produce the tables for the results associated with the double-selected LASSO specification. The result were, fortunately, similar those report in the tables in this section and, hence, conclusions unaltered.

on and by the survey participant by asking after the importance that a 'child works hard' for success in life.

- $family_2$: Marital status.

On the relationship between own employment situation and the effect of the group-level unemployment, Table 6 reports the coefficient on these interaction term defined above.

Table 6: Impact of own employment situation on group-level effects

This table reports results including relevant interaction terms, where the models in the first two-columns corresponds to (5) and (6) in Table 3. Third column corresponds to model (3) in Table (4)

Interaction Effects of Unemployment Status			
	Country (5)	Country (6)	Region (3)
unem_ig	-0.54***	-0.51***	-0.41***
	(0.08)	(0.08)	(0.04)
unem_ig*unemp_country(self reported)	1.55***		
	(0.51)		
Unemp_country(self reported)	-1.87***		
	(0.57)		
unem_ig*Unemp_country(IMF)		1.53**	
		(0.68)	
Unemp_country(IMF)		-2.45***	
		(0.67)	
unem_ig*Unemp_region			0.29
			(0.26)
unem_region			-1.60***
			(0.24)
Observations	74158	74158	38528

Standard errors in parentheses
Standard errors are clustered at country level
The regression is weighted according to population weights
* p<0.10 ,** p<0.05,*** p<0.01

It stands out that for a person who is unemployed, the effect of an increase in the national unemployment rate is associated with a positive increase in the probit adapted life satisfaction measure. This means that when the share of unemployment in a country increases, it is less harmful to be unemployed for those already unemployed. An intuition is that that the individual weights his situation in relation to the general situation in the labor market and that the psychological costs of unemployment decrease though increased social presence of the phenomenon. The interaction effect with the regional level reflects the same picture but is statistically insignificant.

In Table 7 below we report results for assessing the impact of marital status on the effect coming from group-level unemployment.

Table 7: Impacts of marital status on group-level effects

This table reports results including relevant interaction terms, where the models in the first two-columns corresponds to (5) and (6) in Table 3. Third column corresponds to model (3) in Table (4)

Interaction Effects of Married			
	Country (5)	Country (6)	Region (3)
Married(1/0)	0.17***	0.18***	0.16***
	(0.03)	(0.03)	(0.02)
Married(1/0)*unemp_country(self)	0.08		
	(0.13)		
Unemp_country(self)	-1.06***		
	(0.38)		
Married(1/0)*unemp_country(IMF)		-0.10	
		(0.18)	
Unemp_country(IMF)		-2.16**	
		(0.80)	
Married*unemp_region			-0.21
			(0.18)
unem_region			-1.39***
			(0.24)
Observations	74158	74158	38528

Standard errors in parentheses
Standard errors are clustered at country level
The regression is weighted according to population weights
* p<0.10 ,** p<0.05,*** p<0.01

The interaction effect of being married and other people's unemployment status is statistically insignificant for all models and measures of group unemployment levels. The sign of the effect also varies across specification, which leads to the conclusion that married couples happiness responds no differently to aggregate unemployment than non-married's happiness. In

Table 8: Impacts of parental opinions on childrens' work situation on group-level effects

This table reports results including relevant interaction terms, where the models in the first two-columns corresponds to (5) and (6) in Table 3. Third column corresponds to model (3) in Table (4)

Interaction Effects of the Belief, that it is important that children work hard			
	Country (5)	Country (6)	Region (3)
Important that a child works hard(1/0)	-0.28***	-0.20***	-0.17***
	(0.07)	(0.07)	(0.04)
Interaction	0.96*		
	(0.51)		
Unemp_country(self)	-1.51**		
	(0.61)		
Interaction		0.36	
		(0.48)	
Unemp_country(IMF)		-2.17**	
		(0.85)	
Interaction			-0.59**
			(0.26)
Unemp_region			-0.99***
			(0.22)
Observations	74158	74158	38528

Standard errors in parentheses
Standard errors are clustered at country level
The regression is weighted according to population weights
* p<0.10 ,** p<0.05,*** p<0.01

Lastly, as a measure of the attitude towards work within a family, we employ the survey question 'Do you think it is important that children work hard?', where those who reply yes are assumed to have stricter work ethics, not only towards their children, but also towards themselves. Generally, people who agree with this statement have a lower level of overall well-being.

D. Follow-up question 2: The effect of regional versus national unemployment

To investigate if the effect of regional unemployment is the same as national employment, we conducted a range of regressions similar to Table 4, but with the difference that unemployment measured from the IMF database is included in the model alongside the regional-level aggregate. Using an F-test, we test if the regression coefficients between the two unemployment measures (also defining different reference groups) differ significantly. For the regression equivalent to coloumn (1) in Table 4, the effect of a rise in the average national unemployment level affects the individual utility significantly more, then a regional effect. For the regression corresponding to coloumn (3) where we introduce macroeconomic variables but exclude personel effects, there is no significant difference between regional and national effect of unemployment on well being. Lastly we consider column (5) and where we have both macroeconomic and personal characteristics. Here we do again find no significant difference between national and regional effects. It might be the case that both reference groups are so large (the regional and national level), that an individual cannot clearly distinguish the two effects from another. However, we do notice that in absolute terms, the national unemployment rate is higher that the regional for all of the three models considered.

V. Discussion and causal interpretation

Causal interpretation of our estimates has to be undertaken with great caution. The inclusion of a large battery of fixed effect and the LASSO covariate-choice procedure should alleviate endogeneity concerns. Still, most classical endogeneity problems arise. First, it cannot plausibly be ruled out that agents sort between labour markets and regions, and, in particular, sort on unobservables into relatively happy peer groups in thriving labour markets. From this perspective, peer group assignment is not exogenous. One possible strategy to test for sorting between labour markets is to use the hometown at age 14 information given in the European Value Survey, which, with some generosity, could be treated as exogenous, and to specify regression models that control for adolescence home town (unemployment) variables. Strong differences between multiplier coefficients of the presented models and this approach would speak against our implicit assumption of no sorting on unobservables. Second, the problem of simultaneity arises – overall happiness affects individual happiness and vice versa. Without exogenous variation in the data (or theoretically motivated exclusion restrictions), it is impossible to make a statement about the direction of causation. Still, these concerns do not threaten our main result of large and significant happiness spillovers of unemployment on the workers respective reference group.

VI. Conclusion

Unemployment is related to happiness through large negative spillovers on the life satisfaction of peers. Defining reference groups on both the national and the regional level, we estimate unemployment multiplier effects on life satisfaction ranging from 4.5 to 7. The lower bound of these estimators is in line with the previous literature, while the upper estimates are higher than conventional estimates. Our highest estimates occur in the specifications that were estimated on IMF-collected unemployment statistics instead, replacing the European Value survey data, and are always one magnitude larger than previous estimates.

In a second step, we can confirm this finding by implementing a LASSO double-selection procedure, which helps us to choose correct control variables from the micro and macro level and hence minimises the risk of omitted variable bias and variance inflation through multicollinearity. Accordingly, accounting the costs of unemployment only through direct costs on those experiencing personal unemployment is short-sighted. Our results imply that policy makers should be aware of the substantive negative spillover effects of unemployment on life satisfaction.

References

AZUR, M. K., E. A. STUART, C. FRANGAKIS, AND P. J. LEAF (2011): "Multiple Imputation by Chained Equations: What is it and how does it work?" *International Journal of Physciatric Res.*, 20, 40–49.

BELLONI, A., C. D. C. V. AND C. HANSEN (2012): "Sparse models and methods for optimal instruments with an application to eminent domain," *Econometrica*, 80.

BELLONI, A., C. V. F.-V. I. AND C. HANSEN (2017): "Program evaluation and causal inference in high-dimensional data," *Econometrica*, 85, 233–298.

BELLONI A., CHERNOZHUKOV, V. AND C. HANSEN (2014): "Inference on treatment effects after selection among high-dimensional controls," *Review of Economic Studies*.

CLARK, A., A. KNABE, AND S. RATZEL (2010): "Boon or bane? Others' unemployment, well-being and job insecurity," *Labour Economics*, 17, 52–61.

DI TELLA, R. MACCULLOCH, R. J. AND A. J. OSWALD (2003): "The macroeconomics of happiness," *Review of Economics and Statistics*, 85, 809–827.

FREY, B. S. (2008): *Happiness: A revolution in economics:*, Massachusetts Institute of Technology.

GREEN, F. (2011): "Unpacking the misery multiplier: How employability modifies the impacts of unemployment and job insecurity on life satisfaction and mental health," *Journal of Health Economics*, 30, 265–276.

VAN PRAAG, B. AND A. FERRER-I CARBONELL (2004): *Happiness quantified: A satisfaction calculus approach*, Oxford University Press.

VAN PRAAG, B. M., A. FERRER-I CARBONELL, ET AL. (2011): "Happiness economics: A new road to measuring and comparing happiness," *Foundations and Trends® in Microeconomics*, 6, 1–97.

A. Appendix

A. Double-selection procedure using Lasso for control determination

In this section we present the double-selection procedure for estimation the model in (7), providing the Lasso-determined set of relevant controls to include. We define β and γ as targeted variables and are interested in doing inferences on them (and their ratio providing the multiplier effect). We may write the model in general, to fix ideas, as

$$U_{ig} = d_i'\alpha + f(z_{ig}) + \epsilon_{ig}$$
$$d_i = g(z_{ig}) + v_{ig}$$

where $\mathbb{E}\left[\epsilon_{ig}|d_i, x_{ig}\right] = 0$ and $\mathbb{E}\left[v_{ig}|x_{ig}\right] = 0$, $d_{ig} = (u_{ig}, u_g)'$ and is a vector that contains the target variables that we are interested in conducting inference on, x_i is a vector of confounding variables, and ϵ_i and v_i are disturbances. The functions f and g are unknown functions that might be highly non-linear. In order to make inference about $\alpha = (\beta, \gamma)'$ feasible, we suppose that these functions can be approximated by linear function of the control variables $x_{ig} = P(z_{ig})$. To be more precise, we assume that

$$U_{ig} = d_i'\alpha + x_{ig}'\kappa_1 + r_{1ig} + \epsilon_{ig}$$
$$d_{ig} = x_i'\kappa_2 + r_{2ig} + v_{ig}$$

In order to make valid inference about α_0 we have to control in a sufficient manner for the effect of the confounding variables. This is done by allowing x_i to be of a relatively large dimension compared to the sample size, which makes the model described above very flexible. To be more precise, the method works even when the dimension of x_i, denoted by p, is as large as $\log(p) = o(n^{1/3})$. In particular it is possible to include far more variables than what is possible when more traditional methods such as OLS are used. Without further structure it is, however, impossible to conduct inference in the model described above. The key assumption for making inference possible is that of sparsity. Sparsity is the assumption that there exists an approximation $x_{ig}'\kappa_1$ ($x_i g'\kappa_2$) of $f(z_i g)$ ($g(z_i g)$) such that only a small number of the elements of κ_1 and κ_2 are non-zero while the approximation errors r_{1ig} and r_{2ig} are small relative to the estimation error. Formally, we will assume that there exists κ_1 and κ_2 such

$$||\kappa_1||_0 \leq s_n, \quad \text{and} \quad ||\kappa_2||_0 \leq s_n$$

where $s_n \ll n$, while

$$\left(\mathbb{E}\mathbb{E}_n\left[r_{1ig}\right]\right)^{1/2} \leq c\sqrt{s/n}$$
$$\left(\mathbb{E}\mathbb{E}_n\left[r_{2ig}\right]\right)^{1/2} \leq c\sqrt{s/n}$$

for some positive constant c.[13] As mentioned above the high-dimensional-sparse-model framework outlined above extends the standard framework in the literature in two important ways. Firstly, in the traditional framework it is assumed that the number of confounding variables s is small relative to the sample size. This is necessary in order for the standard errors to blow up. Therefore, it is traditionally assumed that the identities of the relevant controls are known, and inference is thus conducted conditional on the assumption of correct model selection. The second contribution is to eliminate the need for this assumption. Instead, we make the far less restrictive assumption that at most s out of a large set of p potential controls can be used to make a sufficiently good approximation of the unknown functions $f(\cdot)$ and $g(\cdot)$. Furthermore, the identity of these controls is not assumed to be known a priori. The sparsity assumption allows us to use the double selection method developed in Belloni A. and Hansen (2014) and Belloni and Hansen (2017) to select approximately the right set of controls and then estimate the effect of the *non-need* variables in d_{ig}. The double selection method works as follows:

1. For each $j = 1,...,d$ run the feasible LASSO using the data $(d_{j,ig}, x_i)$ for $i = 1,...,n$. Let the set of selected variables obtained from doing so be denoted by $\hat{I}_j$.

2. Run the feasible LASSO using the data (U_{ig}, x_{ig}) for $i = 1,...,n$. Let the set of selected variables obtained from doing so be denoted by $\hat{I}_{d+1}$.

3. Let $\hat{I} \supseteq \cup_{j=1}^{d+1} \hat{I}_j$ denote the set of variables that are to be included. Run OLS

[13] Here $\mathbb{E}_n$ denotes the empirical average, and $||\cdot||_0$ is the ℓ_0 norm.

of y on the variables d and the selected elements of x. Traditional (robust) standard errors can be used to make inference on α.

Note that the set $\hat{I}$ may contain more variables that those selected in step 1 and 2 above, however, the sparsity assumption must still be satisfied, i.e. we cannot have $||\hat{I}||_0 > n$. These "extra" variables could be included for additional robustness, however, it would come at the cost of increased standard errors. A particularly nice feature of this approach is that after the model has been refitted (step 3) traditional inference methods can be used. Furthermore, this inference is valid uniformly across a large class of models, see Belloni A. and Hansen (2014) for additional details. Intuitively, this procedure works well not only because it reduced the probability of excluding a relevant control variable, which leads to an omitted variable bias in the estimated effect of the *non-need* variable of interest, but also because it only reduces the probability of including irrelevant control variables which inflates standard errors.

A.1. The feasible LASSO

In this section we briefly describe the feasible LASSO of Belloni and Hansen (2012) which is used in the procedure outlined above. We keep the presentation in the a simplified framework to keep notation simple, but the principles are directly applicable to our setup. Consider a general data set (U_{ig}, x_{ig}), where x is a p dimensional set of control variables. We are interested in applying the feasible LASSO to the following model

$$U_{ig} = x'_{ig}\beta + \xi_{ig}$$

The feasible LASSO is then obtained as the solution to the following penalized quadratic minimization problem

$$\hat{\beta} = \arg\min_{\beta} \mathbb{E}_n \left[\left(U_{ig} - x'_{ig}\beta \right)^2 \right] + \frac{\lambda}{n} ||\hat{\psi}\beta||_1$$

where $\hat{\psi} = diag(\hat{l}_1, ..., \hat{l}_p)$ is a penalty loading matrix and $||\hat{\psi}\beta||_1 = \sum_{j=1}^{p} |\hat{l}_j \beta_j|$. Following Belloni A. and Hansen (2014) the penalty level λ and the penalty loadings l_j for $j = 1,...,p$ are set as

$$\lambda = 2c\sqrt{n}\Phi^{-1}\left(1 - \frac{\gamma}{2p}\right)\hat{l}_j = l_j + o_p(1), \quad l_j = \sqrt{\mathbb{E}(x_{ij}^2 \xi_i^2)}$$

uniformly in $j = 1,...,p$. Here $c > 1$ is a constant to be chosen by the researcher, and $1-\gamma$ is a confidence level. The non-differentiability of the objective function of the LASSO induces the solution, $\hat{\beta}$, to have elements that are exactly equal to zero, thus leading to the desired variable selection. We refer the reader to Belloni A. and Hansen (2014) for a detailed description of how the penalty loadings are estimated. Finally, we note that γ has to be chosen such that $\gamma = o(1)$ and $\log(1/\gamma) \leq K\log(p \vee n)$ for some constant $K > 0$. In practice we will follow the suggestion of Belloni A. and Hansen (2014) and set $c = 1.1$ and $\gamma = (1/n) \wedge 0.05$.