Replication of two studies investigating adaptation in panel data

Anne de Hond

July 2017

1 Introduction

In this document I present my results for the replication of two studies investigating adaptation. I will be applying their methodologies to the SHARE dataset and compare my results [1]. The first study regards adaptation to poverty conducted by Clark, D'Ambrosio and Ghislandi (2016) and the second study by Cubí-Mollà, Jofre-Boner and Serra-Sastre (2016) is concerned with adaptation to chronic illness. As far as the methodology is concerned, the model of Clark, D'Ambrosio and Ghislandi will be easier to implement with my data. The adaptation process studied is different, though, and the results will therefore be harder to compare. On the other hand, the model proposed by Cubí-Mollà, Jofre-Boner and Serra-Sastre is technically more challenging and I ran into some complications whilst trying to apply it to my data. I will discuss the limitations of the studies and my ideas for improvement in the discussion section.

2 Study by Clark, D'Ambrosio and Ghislandi

In a paper by Clark, D'Ambrosio and Ghislandi (2016) adaptation to poverty is analysed. The authors use panel data on almost 54000 individuals in Germany regarding the period from 1985 to 2012. The authors adopt a fixed effects model. They find that life satisfaction is negatively related to both the incidence of poverty and its intensity in a contamporaneous analysis. They hardly find any evidence that the impoverished people adapt to their situation.

[1] DOIs: 10.6103/SHARE.w1.600, 10.6103/SHARE.w2.600, 10.6103/SHARE.w4.600, 10.6103/SHARE.w5.600, 10.6103/SHARE.w6.600, see Börsch-Supan et al. (2013) for methodological details.

2.1 Replication model

For my replication of the Clark, D'Ambrosio and Ghislandi study, I choose to follow their model, albeit changing some of the variables due to our differing subject matter. Their proposed model is as follows:

$$y_{it} = \alpha_i + \gamma_t + \beta C_{it} + \theta I_{it} + \varepsilon_{it}, \tag{1}$$

with y_{it} the self-reported life satisfaction, C_{it} a matrix with covariates and I_{it} a poverty measure at the individual level. As covariates they include age, marital status, labour-force status, residency in East or West Germany, education, number of children in the household and wave dummies.

My model differs from that of Clark, D'Ambrosio and Ghislandi in that y_{it} represents either life satisfaction [2], obtained by the question: "*On a scale from 0 to 10 where 0 means completely dissatisfied and 10 means completely satisfied, how satisfied are you with your life?*" or self-perceived health on a 5-point scale. Note that the lower end of the scale for self-perceived health corresponds to high levels of health. (This means, for example, that we expect that chronic illness will have a positive effect on the self-perceived health scale. I will however refer to this as chronic illness having a negative effect on self-perceived health.) This might be counterintuitive when reading the upcoming results and I will try to change this in future analyses to aid the interpretation. Moreover, I_{it} is an indication of chronic disease or disability, not poverty. The covariates I have included are age, marital status, labour-force status, years of education and number of children. They are converted to resemble to covariates used by the authors. The reference categories for the first four covariates are aged below 51, married, employed and high school education respectively.

In their study on adaptation to poverty, Clark, D'Ambrosio and Ghislandi include all adult respondents with valid information on income and life satisfaction. For my replication study, in general, I select all individuals with valid information on either the life satisfaction question or self-perceived health.

2.2 Replication results

In line with Clark, D'Ambrosio and Ghislandi's approach, I will start with studying the contemporaneous effect of chronic disease on well being. A random sample was drawn consisting of 10000 individuals. This was done to improve the speed of the analysis [3]. I consider d^0 as the incidence of disease and d^1 as the intensity. For d^0 I both consider the incidence of chronic illness and the incidence of a self-reported limitation in instrumental activities of daily living that is larger than 0 (IADL). For d^1, IADL is used.

The results can be found in table 1. In the case of the regression on life satisfaction, the widowed are significantly less satisfied than the married respondents and

[2] Since life satisfaction is only administered in this manner for waves 2 and up, the analysis with life satisfaction is restricted to waves 2, 4, 5 and 6.

[3] I believe my routing is not done very efficiently and it is this that is slowing the analysis down, not the fixed effects regression itself.

the unemployed less satisfied than the employed. For the regression on self-reported health, all age categories of 61+ seem to have a poorer self-reported health on average compared to the reference category of $Age < 51$. Moreover, all employment status categories seem to report a worse self-reported health compared to the employed. Even though Clark, D'Ambrosio and Ghislandi find more significant results, the direction of the effects of the covariates seems to be in general agreement. I find that being chronically ill, whether this is measured through the chronically ill measure or IADL is always significantly and negatively related to self-reported health or life satisfaction. The same goes for the intensity measured with IADL.

Table 1: FE contamporaneous regression

	Life satisfaction	Self-reported health	Life satisfaction	Self-reported health
Age: 51-60	-0.068	0.133	-0.069	0.141
	(0.145)	(0.072)	(0.145)	(0.074)
Age: 61-70	0.116	0.199**	0.110	0.221**
	(0.155)	(0.077)	(0.155)	(0.079)
Age: 71-80	0.271	0.275***	0.265	0.307***
	(0.166)	(0.082)	(0.166)	(0.084)
Age: 80+	0.349	0.378***	0.340	0.413***
	(0.183)	(0.091)	(0.183)	(0.093)
Separated	0.001	-0.102	0.024	-0.179
	(0.321)	(0.160)	(0.321)	(0.164)
Single	0.057	0.197	0.050	0.177
	(0.410)	(0.204)	(0.410)	(0.209)
Divorced	-0.095	0.028	-0.086	-0.001
	(0.215)	(0.107)	(0.215)	(0.110)
Widowed	-0.325**	-0.020	-0.328**	-0.012
	(0.110)	(0.055)	(0.110)	(0.056)
Retired	-0.082	0.118***	-0.081	0.127***
	(0.063)	(0.031)	(0.063)	(0.032)
Unemployed	-0.585***	0.186***	-0.587***	0.202***
	(0.103)	(0.052)	(0.103)	(0.053)
Inactive	-0.076	0.145***	-0.075	0.153***
	(0.076)	(0.038)	(0.076)	(0.039)
Educ < high school	-0.143	0.112*	-0.132	0.098
	(0.101)	(0.050)	(0.101)	(0.051)
Educ > high school	0.020	0.038	0.019	0.055
	(0.100)	(0.049)	(0.099)	(0.051)
No. of children	-0.012	0.020	-0.013	0.025
	(0.030)	(0.015)	(0.030)	(0.015)
Chronically ill	-0.141***	0.395***	.	.
	(0.036)	(0.018)	.	.
IADL	-0.136***	0.096***	-0.083***	0.060***
	0.017	(0.009)	(0.022)	(0.011)
IADL > 0	.	.	-.242***	0.208***
	.	.	(0.061)	(0.031)

Here, the significant codes are as follows: *** indicates $p < 0.001$, ** indicates $p < 0.01$, * indicates $p < 0.05$

Next, in order to study the time profile of the self-reported well being, Clark,

D'Ambrosio and Ghislandi divide the incidence measure d^0 into duration dummies in order to measure the effect duration might have on self-reported well being. They restrict their sample to those individuals for whom the entry into poverty is known. Moreover, only the first spell is taken into consideration. Accordingly, I will retain the observations corresponding to the individuals whose onset of chronic illness is observed and only consider the consecutive time periods in which they persist to have the chronic condition. The results of this regression can be found in table 2 and 3. For brevity's sake the covariates have been omitted.

For the regression with chronic illness as the measure of incidence of chronic disease, we see that the dummies for the regression on life satisfaction become less negative over time and that the dummies for a duration of 2-3 waves and 3-4 waves are not significantly different from 0. This could be an indication of adaptation. However, when we look at self-reported health, there is no obvious change in either direction or the magnitude of the dummy coefficients and all coefficients are still negatively related to self-reported health.

If we look at the results for the regression with ADL as the measure of incidence of chronic disease, we see that only the effect for 2-3 waves on life satisfaction is significantly different from 0 and that it is positive. Moreover, for the regression on self-reported health, the coefficients indicate a somewhat decreasing negative effect of duration on self-reported health with a coefficient that is not significantly different from 0 for a duration of 3-4 waves. Hence, these results hint at the presence of adaptation, although they are not conclusive in and of themselves.

Table 2: FE regression with duration on the basis of chronically ill

	Life satisfaction	Self-reported health
Chronically ill 0-1 waves	-0.151***	0.428***
	(0.016)	(0.008)
Chronically ill 1-2 waves	-0.077**	0.468***
	(0.029)	(0.015)
Chronically ill 2-3 waves	-0.069	0.462***
	(0.051)	(0.051)
Chronically ill 3-4 waves	0.022	0.467***
	(0.099)	(0.051)
IADL	-0.220***	0.120***
	(0.012)	(0.005)

Here, the significant codes are as follows: *** indicates $p < 0.001$, ** indicates $p < 0.01$, * indicates $p < 0.05$

Table 3: FE regression with duration on the basis of IADL incidence

	Life satisfaction	Self-reported health
IADL > 0 0-1 waves	-0.060	0.185***
	(0.031)	(0.014)
IADL > 0 1-2 waves	0.036	0.190***
	(0.054)	(0.025)
IADL > 0 2-3 waves	0.277**	0.166***
	(0.097)	(0.044)
IADL > 0 3-4 waves	0.130	0.032
	(0.201)	(0.091)
IADL	-0.163***	0.059***
	(0.013)	(0.006)

Here, the significant codes are as follows: *** indicates $p < 0.001$, ** indicates $p < 0.01$, * indicates $p < 0.05$

3 Study by Cubí-Mollà, Jofre-Boner and Serra-Sastre

The study by Cubí-Mollà, Jofre-Boner and Serra-Sastre (2016) focuses on adaptation to a long-standing illness (LSI) via measuring changes in self-assessed health. The data comes from the British Cohort Study and consists of four waves from 1970 onwards. The authors apply a dynamic ordered probit model. Their results support the existence of adaptation.

3.1 Replication model

Cubí-Mollà, Jofre-Boner and Serra-Sastre select individuals reporting to have only one LSI under the assumption that those will keep their underlying health constant.

Their empirical strategy utilizes a latent health model according to:

$$SAH_{it}^{*} = \alpha SAH_{i,t-1} + \beta m_{it} + \delta d_{it} + \boldsymbol{\gamma'} \boldsymbol{x_{it}} + c_{it} + u_{it}, \tag{2}$$

where SAH_{it} and $SAH_{i,t-1}$ represent self-assessed health in periods t and $t-1$. Moreover, m_{it} stands for morbidity corresponding to individuals having **one** chronic condition and d_{it} measures the duration since the onset of illness. As additional explanatory variables, x_{it}, they include number of children in the household, marital status, employment status, tenure regarding the housing situation and education received.

Cubí-Mollà, Jofre-Boner and Serra-Sastre model the latent health by means of an ordered probit model, with

$$SAH_{it} = k \text{ if } \lambda_{k-1} < SAH_{it}^{*} < \lambda_k \text{ for } k = 1, \ldots, K. \tag{3}$$

Here, K is the total number of categories and $\lambda_0 = -\infty$ and $\lambda_K = \infty$. The probability for individual i in period t of reporting a specific SAH category becomes:

$$\begin{aligned} P(SAH_{it} = k) = \Phi(\lambda_k - \alpha SAH_{i,t-1} - \beta m_{it} - \delta d_{it} - \boldsymbol{\gamma'} \boldsymbol{x_{it}} - c_i) \\ - \Phi(\lambda_{k-1} - \alpha SAH_{i,t-1} - \beta m_{it} - \delta d_{it} - \boldsymbol{\gamma'} \boldsymbol{x_{it}} - c_i), \end{aligned} \tag{4}$$

with $\Phi(.)$ the standard normal cumulative distribution function.

In this nonlinear fixed effects model, the unobserved heterogeneity presented by c_i cannot simply be removed by applying a linear transformation. In fact, no simple transformation is known to exist that eliminates these terms in a nonlinear setting. Hence, the incidental parameter problem arises, presenting us with $N+k$ parameters to be estimated where k is the number of explanatory variables on the right side of the equation. Moreover, the dynamic aspect of this specification can lead to inconsistent estimators if regular estimation techniques are applied. The authors suggest dealing with the dynamic and incidental parameter problem by using Wooldridge's (2005) approach, which suggests the parameterization of the fixed effects c_i as a function of the first observed SAH in the sample and the average of the exogenous variables $\bar{\boldsymbol{x}}_{\boldsymbol{i}}$:

$$c_i = \sigma + \phi SAH_{i,1} + \mu\bar{m}_i + \nu\bar{d}_i + \boldsymbol{\kappa'}\bar{\boldsymbol{x}}_{\boldsymbol{i}} + \varepsilon_i \tag{5}$$

Note that the correct specification of the parameterization of c_i is crucial to the consistency of the estimates. The final equation can be obtained by substituting the results in equation 5 back into equation 4 producing:

$$\begin{aligned} P(SAH_{it} = k) = \Phi(\lambda_k - \alpha SAH_{i,t-1} - \beta m_{it} - \delta d_{it} - \boldsymbol{\gamma'}\boldsymbol{x}_{\boldsymbol{it}} - \sigma - \phi SAH_{i,1} - \mu\bar{m}_i - \nu\bar{d}_i - \boldsymbol{\kappa'}\bar{\boldsymbol{x}}_{\boldsymbol{i}}) \\ -\Phi(\lambda_{k-1} - \alpha SAH_{i,t-1} - \beta m_{it} - \delta d_{it} - \boldsymbol{\gamma'}\boldsymbol{x}_{\boldsymbol{it}} - \sigma - \phi SAH_{i,1} - \mu\bar{m}_i - \nu\bar{d}_i - \boldsymbol{\kappa'}\bar{\boldsymbol{x}}_{\boldsymbol{i}}), \end{aligned} \tag{6}$$

In my replication of the results of Cubí-Mollà, Jofre-Boner and Serra-Sastre I aim to follow their procedure as closely as possible. Hence, I only select the individuals with one chronic disease throughout all the SHARE waves. This leaves me with 4470 individuals and 8935 observations. This is notably less than the 11493 observations used by Cubí-Mollà, Jofre-Boner and Serra-Sastre.

For the dependent and lagged variable SAH I use my measure of self-perceived health which is recorded in the exact same way as that of the authors on a five point scale. An important difference in interpretation is that 1 on my scale corresponds to *Excellent* and 5 to *Poor*, thereby reversing the interpretation of the scale compared to that of Cubí-Mollà, Jofre-Boner and Serra-Sastre, but not altering the estimation. In accordance to their study, I collapse the *Excellent* and *Very good* categories, leaving me with four final categories [4]. In my analysis, m_{it} is a dummy variable corresponding to individual i having one chronic disease in period t. The duration variable had to be generated by me [5]. I followed the authors' suggestion in taking the average time between two waves as an indication of duration when a respondent first reports to have a chronic illness in the most recent of the two waves. The sociodemographic variables included were chosen to concur with those of the authors and were converted in a similar manner. The reference categories for SAH, marital status, employment status, tenure and education are an SAH of poor, being single, being employed, another type of tenure and no education qualifications.

[4] In the original study by Cubí-Mollà, Jofre-Boner and Serra-Sastre, the question on self-assessed health changes the number of categories across waves. According to the authors, collapsing categories does not affect the estimations of covariates.

[5] There is a variable available for this in the SHARE data and for future analysis this is probably the better choice.

3.2 Replication results

The results of the estimated coefficients of the ordered probit model can be found in table 4. For brevity's sake I only report the estimates for the coefficients of the lagged and the first sample period of the SAH, the incidence of having one chronic illness (LSI) and the duration. In agreement with Cubí-Mollà, Jofre-Boner and Serra-Sastre, I find a strong state dependence apparent in the positive coefficients in the first column of table 4. Moreover, the positive coefficients on $SAH_{i,1}$ clearly indicate that the initial period's SAH determines SAH in consecutive periods.

The second column presents the coefficients corresponding to a model where the incidence of having one chronic disease is added. Unlike Cubí-Mollà, Jofre-Boner and Serra-Sastre, I do not find that adding this variable absorbs part of the effect that the previous health states have on the current SAH. The coefficent on LSI is positive, indicating that having a chronic disease lowers somebody's health state evaluation. In the last column of table 4, duration has also been added. Contrary to what Cubí-Mollà, Jofre-Boner and Serra-Sastre results, I do not find any evidence of adaptation, with the coefficient on the duration variable being significant and positive, albeit very small.

Table 4: Ordered probit model

	Dynamic model	Dynamic model with incidence	Dynamic model with incidence and duration
$SAH_{t-1} = Fair$	0.481*	0.508**	0.518**
	(0.216)	(0.218)	(0.218)
$SAH_{t-1} = Good$	1.040***	1.069***	1.134***
	(0.216)	(0.218)	(0.218)
$SAH_{t-1} = Excellent$	1.601***	1.662***	1.760***
	(0.218)	(0.220)	(0.221)
$SAH_{t,1} = Fair$	1.042***	1.114***	1.038***
	(0.237)	(0.239)	(0.239)
$SAH_{t,1} = Good$	1.792***	1.898***	1.924***
	(0.239)	(0.241)	(0.241)
$SAH_{t,1} = Excellent$	2.513***	2.678***	2.759***
	(0.241)	(0.242)	(0.243)
LSI	.	0.489***	0.379***
	.	(0.035)	(0.042)
Duration	.	.	0.074***
	.	.	(0.014)
Threshold 1	-3.768	-3.274	-3.241
Threshold 2	-2.051	-1.506	-1.457
Threshold 3	-0.602	-0.019	0.048

Standard errors are obtained via the inverse of the negative hessian calculated with a maximum likelihood approach at the optimized estimates. Here, the significant codes are as follows: *** indicates $p < 0.001$, ** indicates $p < 0.01$, * indicates $p < 0.05$

4 Discussion

Considering the replication results of the study by Clark, D'Ambrosio and Ghislandi, I find that there could indeed be adaptation to chronic illness, although my results do not provide sufficient evidence. The coefficients on the duration dummies become increasingly less negatively related to good health and in some model specifications even positive. Hence, it might be worthwhile to study the effect of longer duration spells and see whether this observed trend develops into fully observable adaptation.

I do not find the positive effect of duration on self-reported health that Cubí-Mollà, Jofre-Boner and Serra-Sastre find. The most likely explanation for this is that the authors only find a significant effect of duration after 20 years. Due to the prospective aspect of my sample (I only include individuals whose onset of disease is within the sampling period), I have only duration spells of a maximum of 8.5 years.

In line with both the results mentioned above, I suggest including individuals with a longer duration of chronic illness than those that I have studied thus far. Even though I loose the purely prospective nature of the study - where I can compare the effect of chronic illness within individuals due to the fact that we observe the onset of chronic illness - I gain the possibility to study the long-term effects of duration on chronic illness.

Another possible reason for the lack of adaptation in the SHARE sample is that Cubí-Mollà, Jofre-Boner and Serra-Sastre investigate a fundamentally different sample, with subjects ranging from 26 to 38 years old, whereas the SHARE data is exclusively focused on subjects of 50 years and up. Hence, it could be the case that older individuals simply do not adapt as easily as younger ones. Alternatively, the chronic conditions prevalent in the younger sample might be different to those reported by the older sample and the adaptation process for these subsets of diseases could differ. Since the SHARE database simply does not provide data on younger individuals I cannot perform a comparative analysis for age with the current data and will leave this to future research.

A different cause for the discrepant results could be that the selection procedure proposed by Cubí-Mollà, Jofre-Boner and Serra-Sastre is not appropriate for the SHARE data. First of all because this leaves me with only a very limited number of observations, since the majority of the chronically ill subjects will report to have more than one chronic illness. Second of all, their justification for selecting the subjects reporting to have one long-standing illness is the assumption that the latent health of those individuals will be stationary. However, this does not seem like a very plausible assumption when one also considers diseases like cancer and rheumatoid arthritis, which have been known to be progressive in nature. Hence, it would be insightful to broaden the focus to nonstationary diseases when it comes to studying adaptation, since a broader range of illnesses can be studied and no limiting assumptions regarding latent health will have to be made.

Finally, a disadvantage of the methodology of Cubí-Mollà, Jofre-Boner and Serra-Sastre is that it heavily relies on the correct parameterization of the individual specific effects. It is true that this approach notably simplifies the analysis, but if the parameterization is misspecified the estimations will be inconsistent. Hence, I propose

to use a dichotomized ordered logit model in which no specification for the fixed effects is required and that still produces consistent results. This estimation method is described in Baetschmann, Staub and Winkelmann (2011) and makes use of a conditional likelihood that has been studied for the binary logit model.

REFERENCES

Baetschmann, G., Staub, K.E. & Winkelmann, R. (2015). Consistent estimation of the fixed effects ordered logit model. *Journal of the Royal Statistical Society, 178(3),* 685-703.

Börsch-Supan, A., Brandt, M., Hunkler, C., Kneip, T., Korbmacher, J., Malter, F., Schaan, B., Stuck, S., Zuber, S.; SHARE Central Coordination Team. (2013). Data resource profile: The Survey of Health, Ageing and Retirement in Europe (SHARE). *Int J Epidemiol., 42(4),* 992-1001.

Clark, A.E., D'Ambrosio, C. & Ghislandi, S. (2016). Adaptation to poverty in long-run panel data. *Review of Economics and Statistics, 98(3),* 591-600.

Cubí-Mollà, P., Jofre-Bonet, M. & Serra-Sastre, V. (2016). Adaptation to health states: Sick yet better off? *Health Economics,* 1-18.

Wooldridge, J.M. (2005). Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. *Journal of Applied Econometrics, 20,* 39-54.