Comment on "Unemployment Compensation and Wages: Evidence from the German Hartz Reforms" by Stefan Arent and Wolfgang Nagl

Ludsteck, Johannes and Seth, Stefan

IAB Nürnberg (Institute for Employment Research Nuremberg)

21 October 2013

Comment on "Unemployment Compensation and Wages: Evidence from the German Hartz Reforms" by Stefan Arent and Wolfgang Nagl

Johannes Ludsteck and Stefan Seth*
Institute for Employment Research (IAB) Nürnberg

October 21, 2013

Abstract

Arent and Nagl (2013) use the BA Employment panel 1998-2007 to identify effects of the German Hartz reform. Their findings suggest that the reform caused a considerable reduction of wages. Our replication of their study suggests that their clear and strong conclusions are driven by a too coarse modelling of the time effects. They become blurred and weak once the development of wages is investigated based on a finer time grid. Further methodological considerations put into question whether their models are well-suited to investigate the wage effects of the reform.

JEL classification: J08; J31; J65
Keywords: Hartz reforms; unemployment compensation; wages

1 Introduction

In an interesting and ambitious paper, Stefan Arent and Wolfgang Nagl try to exploit the German Hartz reform as a natural experiment in order to evaluate the effect of unemployment compensation on wages. The authors estimate fixed effects regressions of log wages on individual, firm, and industry level control variables. They try to capture the effect of a decrease of unemployment benefits by the coefficient of a (post-reform) period dummy for the years 2005 to 2007. They interpret their results as "strong evidence that decreased unemployment compensation has an adverse effect on wages."

We extend their regression models slightly by replacing the post-reform period dummy by a full set of year dummies. This yields the time series of wages, purged from other influences. If this finer grid (year by year) is used to investigate the development of the purged wages, we find that wage decreases took place already in 2004. The further decreases in the years 2005 to 2007

*We are greatly indebted Ekkehart Schlicht and Helmut Rudolph for helpful suggestions and comments. We thank Stefan Arent and Wolfgang Nagl for sharing their data preprocessing scripts with us. Their openness and willingness to cooperate facilitated our work considerably. Of course, all remaining errors are our own.

follow a smooth trend but show no clear jump. Since the integration of the unemployment insurance with the social assistance benefits took place in 2005 and the reduction of the maximum unemployment entitlement period became effective in February 2006, it appears daring to give the 'post'-reform dummy (that actually includes the period **2005** to 2007) a causal interpretation.

2 Arent and Nagl's Model

Arent and Nagl (2013) employ the BA-Emplomynent Panel 1998–2007 (but use only the years 2000–2007 in their regression samples) to estimate wage regressions of the form

$$\ln(w_{it}) = \beta_0 + \beta_1\, LUA_t + controls + a_i + u_{it} \tag{1}$$

where w_{it} denotes real monthly wages of indivual i in quarter t, LUA_t is a period dummy taking on value 1 for the years 2005–2007 (and zero otherwise), *controls* contains individual, establishment and industry level control variables,[1] a_i denotes an individual specific fixed effect, and u_{it} denotes a residual term. They emphasize that the Hartz reform generates quasi-experimental conditions and argue that purging the wages from the control variables in their model allows to interpret the coefficient of the *LUA* dummy ('Lower Unemployment Assistance') as a measure for the effect of lowering the unemployment assistance on wages.

We find it difficult to follow this argument for the following reason. Experimental and quasi-experimental designs usually compare the development of a group affected by a reform (the treatment group) with that of another group (the control group) *that is subject to the same common influences except the reform.* This allows the researcher to ignore influences that affect *both groups in the same way.* Since a control group is absent here (*all* German employees are affected by the reform), controlling for macro level impacts is required. Arent and Nagl are aware of the problem and try to tackle it by including the industry level gross value added per worker as a macro level control. It appears highly problematic as a regressor, however, since it is defined as a function of wages[2] and therefore must be highly correlated with the dependent variable *by construction.* Arent and Nagl also stress that controlling for a number of individual and firm level characteristics helps to create quasi-experimental conditions. Their controls are, however, almost time-invariant (the skill dummies) or change steadily over time (the age profile, firm size dummies and the firm's age structure). Therefore their potential to capture the effects of discrete changes of macro variables (as e.g. export demand indicators) is quite limited.

3 A slight extension of their model

Even if one provisionally accepts Arent and Nagl's assumptions, it becomes difficult to find a pronounced reform effect in 2005 if the reform dummy *LUA*

[1] The controls include age, age squared, professional status, firm size, the firms's age structure, individual job tenure, annual values of the industry-specific gross value added per worker and dummy variables for each quarter.

[2] The GVA is computed as output at market prices minus intermediate consumption. Clearly, wages constitute a large share of the value added.

is replaced by dummies for the years 2002 to 2007.[3] This yields the following model

$$\ln(w_{it}) = \beta_0 + \sum_{\tau=2002}^{2007} \gamma_\tau \, D_t^{(\tau)} + \mathit{controls} + a_i + u_{it} \tag{2}$$

where $D_t^{(\tau)}$ is a year dummy for year τ. It takes on value 1 for all quarters of year τ and zero otherwise. When model (2) is estimated for this sample, the dummy coefficients show a trend pattern that decreased already in 2004, one year before Arent and Nagl located the structural break, and two years before the reduction of the maximum entitlement period became effective. This suggests that their Hartz effect might be an artifact resulting from cramming the time series of the purged wages into a too coarse step function.

4 Results

Arent and Nagl obtain a *LUA* coefficient value of -0.024 (with standard deviation 0.00015) when they estimate model 1 for western German men. According to their interpretation, the Hartz reform reduced wages by 2.4 percent for the western German men. The step function generated by their dummy effect is represented in Figure 1.

Before we present the results from the extended model, we note that the estimated Hartz effect seems to be much less robust than suggested by its tiny standard error (0.0002). If we run separate regressions based on the observations from one quarter only, the reform dummy estimates for the first, second, third and fourth quarter are -0.018, -0.022, -0.027 and -0.029, respectively. Considering that the standard errors of the quarter-specific reform dummy coefficients are still very small (0.0003), and that the dummy coefficient measures an average over the three years 2005, 2006 and 2007, it is quite irritating that the 'Autumn' Hartz effect exceeds the 'Winter' effect by roughly 60 percent. A possible explanation is that a longer estimation period would be required to obtain precise joint estimates of the age profile and the period dummy which measures deviations from this profile.[4] Since our year dummies are also prone to this collinearity problem, our fixed effects Tobit estimates are based on a longer estimation period (covering the period 1995 to 2007). Further thoughts on the error structure of the model suggest that the reported standard errors might be underestimated by an order of magnitude since aggregate shocks that are not represented in the regression model may generate clustering of the residuals at the industry or year level. Accounting for that might enlarge the standard errors considerably.[5]

Anyway, things look quite different when model (2) is estimated instead. To demonstrate this visually, the coefficients of the year dummies $\gamma_{2002}, \ldots, \gamma_{2007}$

[3] The year dummies for 2002 and 2001 are omitted to avoid collinearity problems with the age terms. Deviating from Arent and Nagl (they include the age variable at yearly precision in their models), we use the age variable at quarterly precision in order to avoid errors-in-variables bias.

[4] Note that the fixed effects regressions exploit the *changes* of the regressors to identify the coefficients.

[5] A simple two-step approach (based on Amemiya (1978)) that yields more realistic standard errors would be to regress time series of the dummy coefficients from model 2 on the Hartz dummy and other macro level controls. It yields a standard error of 0.005 for A & N's *LUA* coefficient estimate -0.024.

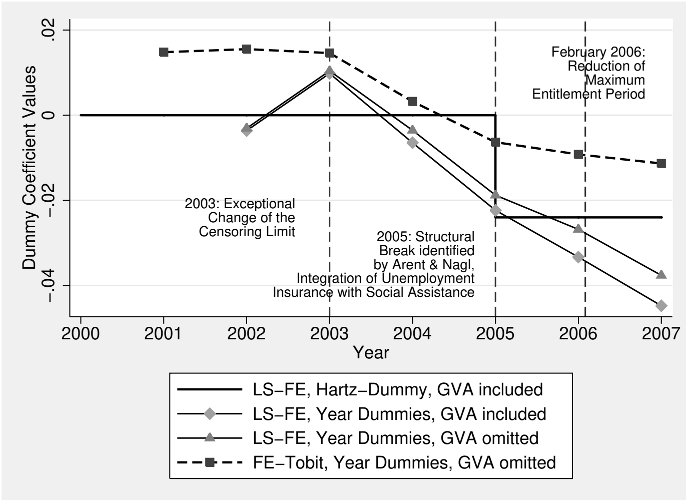

Figure 1: Dummy coefficient estimates, obtained from estimating models (1) and (2) for western German men.

Data Sources: BA-Employment Panel (LS-FE, data preprocessing by Arent and Nagl), and SIAB (FE-Tobit, data preprocessing by Ludsteck and Seth).

Legend: 'LS-FE, Hartz-Dummy' represents Arent and Nagl's original Hartz-Dummy effect from model (1), 'LS-FE, Year Dummies' represents standard least squares fixed effects estimates of the year dummy coefficients for model (2). 'FE-Tobit, Year Dummies' refers to fixed effects Tobit estimates based on a highly similar sample from the SIAB. 'GVA included' and 'GVA omitted' indicate whether the gross value added per worker is included or not as a control variable.

are plotted for Arent and Nagl's first sample (western German men) in Figure 1. Irrespectively of whether the gross value added is or is not included as a control, the dummy coefficients show a peak in 2003 and a steady downard trend starting already in 2004. The graphs show clearly that wages are *cet. par.* smaller in the years 2005–2007 on average, but the changes follow a downward trend and there is no indication of a structural break or a jump in 2005.

Further inspection of the data suggests that the peak in 2003 may be caused by improper treatment of the wage censoring. As is well known, the wage of the employment register data (Beschäftigtenstatistik) are right-censored at the earnings limit for social security contributions ('Beitragsbemessungsgrenze'). The censoring share is moderate for the entire sample, about 5 to 10 percent, slightly varying by year. For the highly qualified men in western Germany, it amounts to roughly 50 percent, however.[6] This may bias standard OLS and fixed effects regression models severely and therefore requires proper treatment, e.g. by using Tobit models.

Arent and Nagl try to tackle the censoring problem heuristically by truncating the sample at the 5th and 95th percentile of the unconditional wage distribution. This is problematic for several reasons. First, the application of standard OLS or FE models to a truncated sample delivers still biased results.

[6]See e.g. Büttner and Rässler (2008) for statistics on the censoring shares.

Since roughly 50 percent of the wages are censored for the highly qualified men in western Germany, comparing the results from standard OLS or fixed effects regressions for this group with the unskilled workers (where censoring is below 2 percent), appears to be problematic, to say the least. Second, Arent and Nagl's heuristic truncation procedure leaves the majority of the censored observations in the sample if the share of right-censored observations exceeds 5 percent, which clearly is the case for several subsamples (e.g. western German men and the highly qualified employees).[7] Third, the censoring limit was increased by roughly 10 percent in 2003.[8] If such changes are not tackled appropriately using Tobit models that allow for variable censoring limits, these changes may translate directly into artificial changes of the year or period dummy coefficients.

We inspect whether the the peak of the dummy coefficient pattern in 2003 is caused by the sizeable increase of the censoring limit by re-estimating model (2) using Honoré's (1993) fixed effects Tobit estimator. This exercise requires some other adjustments. First, Arent and Nagl's data cannot be used due to lack of correctly defined censoring indicators. Second, the year dummy coefficients measure deviations from the (quadratic) age profile. Therefore it appears to be difficult to obtain precise joint estimates of the age profile and the year dummies in short panels. We reduce the problem by extending the estimation period to 1995 and including year dummies only for the years 2001 to 2007. (The dummy coefficients measure then deviations from the average over the years 1995 to 2000). Since the BA-employment panel starts in 1998 and contains all information required for estimation only for the years 2000 to 2007, we employ the SIAB data set to perform the Tobit estimates.[9] The estimation period covers the years 1995 to 2007 and includes year dummies for 2001 to 2007. See the Appendix for a more detailed description of the model and the data preprocessing.

The results from this exercise are represented by the dashed graph in Figure 1. We find again a wage decline that starts already in 2004 and gives no indication of a clear-cut downward kink or jump afterwards. The graph becomes even flatter in the years 2006 and 2007. It is hard to imagine that a researcher who didn't know Arent and Nagl's article and further details on the Hartz reform would find a negative reform effect in the years 2005 to 2007 based on the year dummy coefficients.[10]

[7] Due to an error in their preprocessing scripts, Arent and Nagl actually drop only about 1 percent of the top observations instead of 5 percent as reported in the paper.

[8] The increase was induced by the Beitragssicherungsgesetz (BSSichG), see Bungesgesetzblatt, Jahrgang 2002 Teil I, Nr. 87.

[9] Estimations based on the SIAB yield generally almost the same results as the BA employment panel since both data sets are based on the same source, the employment register. The SIAB is better suited to analyse wages than the BA employment panel since roughly 5 percent of wages are imputed in the latter ('Fortschreibungsfälle').

[10] Arent and Nagl find a structural break in 2005 based on a Chow test (see page 454 of their article). Unfortunately they do not report any further details of the test procedure. When we performed standard Chow tests (by interacting all regressors with the period dummy and testing for joint significance of all interaction effects), we obtained a structural break for every year. Given the high precision of the estimates, this does not come as a surprise. Arent and Nagl try to corroborate their finding by additional pooled regressions based on sub-periods before (2003) and after the reform (2007). It is, however, completely unclear what to learn from these regressions. They are based on *linear* models (see Appendix Tables A1 and A2) whereas the regressions in the main part are *log-linear*. If the true model is log-linear, then the linear one is misspecified and structural breaks may either appear artificially or be hidden as a consequence of neglected nonlinearity.

In principle, Appendix Table A4 in Arent and Nagl's paper contains already enough information to put their interpretation of the Hartz dummy into question. They run placebo regressions to check the validity of their hypothesis. The placebo specifications are obtained by shifting the start of the Hartz period dummy successively to 2004, 2003 and 2002. Whereas the coefficient of the Hartz dummy starting in 2005 is significantly negative (-0.024), shifting the start backwards to 2004 and 2003 and 2002 renders the coefficients positive (see Table A4 in their paper). Their values are -0.008, 0.0266 and 0.0213 for the respective years 2004, 2003 and 2002 (all significant at the 1 percent level). Combining these estimates allows to retrieve the pattern of year-specific deviations from the overall trend and clearly shows that the downward movement started already before 2005. In general, highly significant placebo effects should *not* be taken as evidence of the validity of an experimental design, even if they have the 'wrong sign.' Effective placebos rather indicate the presence of ignored confounding factors.

5 Conclusion

Arent and Nagl have, to their merit, raised an important empirical issue: What was the effect of the Hartz reforms on wage formation? They try to answer it based on simple fixed effects regression models including a post-reform dummy. Due to the absence of a control group and questionable control for confounding factors at the macro level, their ceteris paribus interpretation of the reform effects is based on quite strong and untested assumptions, however. Moreover, slight extensions of their models suggest that part of these wage decreases *preceded* the integration of unemployment security and social assistance benefits in 2005 and the reduction of the maximum entitlement period that became effective in February 2006. Consequently, it appears rather daring to attribute the observed effects to the Hartz reform. Given the problems with Arent and Nagl's analysis, their conclusion remains unconvincing, and the question about the empirical effects of the Hartz reforms remains unsettled.

Any wholesale assertion like "Economic theory predicts that decreasing unemployment compensation leads to decreased wages and higher rates of employment" seems to us to underestimate the flexibility of economic theory considerably, as other outcomes could easily be rationalized in terms of models that emphasize on-the-job search, rather than off-the job-search, for example. With on-the-job search, wage offers are compared to the *prevailing* wage rate, rather than unemployment benefits. Given the rather low unemployment rates of the better skilled workers, on-the job search seems to be of more relevance for them than to the lower-skilled workers. This would account for the increasing wage gap between the better skilled workers and the less skilled workers we observe.[11] In contrast, the view suggested by Arent and Nagl would suggest, counterfactually, a narrowing wage gap between the highly skilled and the less skilled workers in the wake of the Hartz reforms.

[11] See e.g. Card et al. (2013)

References

Amemiya, T. (1978). A note on an Random Coefficients Model. *International Economic Review*, 19(3): 793–796.

Arent, S., and Nagl, W. (2013). Unemployment Compensation and Wages: Evidence from the German Hartz Reforms. *Jahrbücher für Nationalökonomie und Statistik*, 233(4): 450–466.

Büttner, T., and Rässler, S. (2008). Multiple Imputation of Right-Censored Wages in the German IAB Employment Sample Considering Heteroscedasticity. Discussion Paper 44/2008, IAB, Nürnberg.

Card, D., Heining, J., and Kline, P. (2013). Workplace Heterogeneity and the rise of West German Wage Inequality. *Quarterly Journal of Economics*, 128(3): 967–1015.

Honoré, B. (1993). Trimmed LAD and Least Squares Estimation of Truncated and Censored Regression Models with Fixed Effects. *Econometrica*, 60(3): 553–565.

A Details on the data preprocessing and the implemenation of Honoré's fixed effects Tobit estimator

The application of Honoré's estimator to Arent and Nagl's data requires some minor and straightforward adjustments. As we argue in the main section above, using a longer estimation period provides a panorama view on the effects of the Hartz reform and should give more reliable results. To achieve both goals, we employ the SIAB data set.

The following subsection describes the adjustments of the data preprocessing. It is followed by a brief description of the estimation procedure.

A.1 Data preprocessing

The preprocessing of the SIAB data follows Arent and Nagl with some minor exceptions. First, the establishment level shares of old and young workers are not available in the SIAB. Since these shares show only very small time variation, omitting them is completely innocuous, however. (We tested that by re-running A&N's regressions without the shares and found that the time pattern of the year dummies is almost identical to the one obtained from including.) Second, an inspection of the skill dummies in Arent and Nagl's regressions shows that they show extremely low time-variation.[12] The dummies for unskilled blue collar workers, for skilled blue collar workers and for foremen change only in 0.22

[12]It is clear that the skill dummies are almost time-constant since the formal qualification (university degree) is either acquired before persons enter the labour market, or it is acquired in an apprenticeship. In the second case, the change of the qualification is not visible in the estimation sample since then the change occurs at the end of the apprenticeship and apprenticeship spells are removed from the estimation sample.

percent, 0.22 percent and 0.04 percent of all observations in Arent and Nagl's data. These extremely sparse dummies may create optimization problems and are therefore omitted from our Tobit models. For sake of simplicity we omit also the dummies for small an large firms. We checked whether this has an effect on the year dummy coefficients by running model standard least squares fixed effects regressions without these controls and found only tiny (visually almost invisible) effects. Third, we use only spells crossing the 30th of June of every year (i.e. yearly data) to keep the observation numbers within limits.[13] The resulting data set contains 1,602,587 observations.

Our sample selection for the estimation sample 1995 to 2007 follows Arent and Nagl's sample selection on the heels for sake of comparability. We find some of their choices, however, problematic. For example, their balancing approach (they include only persons observed in every quarter between 2000 and 2007) drops all persons entering or leaving the labour market due to natural cohort turnover. They keep also persons older than 65 years in their estimation sample.

A.2 Details on the implementation of Honore's estimator

The application of Honoré's fixed effects Tobit estimator to our models and data require some minor adjustments. These are described here.

First, Honoré's estimator is formulated for time-constant left-hand side censoring at zero only. Fortunately, it can be applied directly after application of the simple transformation $\tilde{y}_{it} := c_t - y_{it}$ to the censored dependent variable. If y_{it} is right-censored at time-varying limits c_t, then $\tilde{y}_{it}$ is left-censored at zero. This (deterministic) transformation switches the signs of all coefficients and shifts the constant and the time dummies. To retrieve the original dummy coefficients γ_τ from the estimates $\tilde{\gamma}_\tau$, we need the transformation $\gamma_\tau := c_\tau - \bar{c}_{[1..6]} - \tilde{\gamma}_\tau$.[14]

Second, the orthogonality conditions used to estimate the coefficients of Honoré's model are formulated for one pair of waves only. To apply the estimator to a sample spanning more that two waves, we simply minimize the sum of the objective functions for several wave-pairs. To represent this formally, define $\tilde{Y}_t := (\tilde{y}_{1t}, \ldots, \tilde{y}_{Nt})'$ and $X_t := (x_{1t}, \ldots, x_{Nt})'$. If $\chi^2(\tilde{Y}_{t-1}, \tilde{Y}_t, \Delta X_t\, b)$ denotes the objective function for the wave pair $(t-1, t)$, the objective for the entire sample has the form

$$\Omega = \sum_{t=2}^{T} \chi^2(\tilde{Y}_{t-1}, \tilde{Y}_t, \Delta X_t\, b). \qquad (3)$$

It is minimized using a derivative-free global optimization algorithm.[15]

[13]Note that roughly 45 percent of all employment spells in the estimation period refer to a whole year ('Jahresspells') and that these spells account for roughly 70 percent of the total working time (in days). Since the Jahresspells contain the 12-month average of wages, using quarterly data does not add much information since then *identical* wage information is repeated four times for the 'Jahresspells.'

[14]The mean of the censoring limits for the base period $\bar{c}_{[1..6]} \equiv \sum_{t=1}^{6} c_t/6$ has to be subtracted from c_t since the dummy coefficients represent deviations from the basis period $t = 1, \ldots, 6$ corresponding to the years $1995, \ldots, 2000$.

[15]We use Nelder and Mead's flexible polyhedron search as implemented in the software package *Mathematica*. Since the optimization procedure showed starting-value dependence, the algorithm was restarted 50 times with random starting values in the range $[-0.1; 0.1]$ in order to increase the likelihood to find the global minimum. The program code is available from the authors upon request.