IAW DISCUSSION PAPERS

iaw@iaw.edu | www.iaw.edu

IAW Diskussionspapiere Nr. 135 Januar 2021 | No. 135 January 2021

Wage expectation, information and the decision to become a nurse

Philipp Kugler

Institut für Angewandte Wirtschaftsforschung e.V.
Schaffhausenstraße 73 | 72072 Tübingen | Germany
Tel.: +49 7071 98960 | Fax: +49 7071 989699

ISSN: 1617-5654

IAW-Diskussionspapiere

Das Institut für Angewandte Wirtschaftsforschung (IAW) Tübingen ist ein unabhängiges außeruniversitäres Forschungsinstitut, das am 17. Juli 1957 auf Initiative von Professor Dr. Hans Peter gegründet wurde. Es hat die Aufgabe, Forschungsergebnisse aus dem Gebiet der Wirtschafts- und Sozialwissenschaften auf Fragen der Wirtschaft anzuwenden. Die Tätigkeit des Instituts konzentriert sich auf empirische Wirtschaftsforschung und Politikberatung.

Dieses IAW-Diskussionspapier können Sie auch von unserer IAW-Website als pdf-Datei herunterladen:

http://www.iaw.edu/index.php/IAW-Diskussionspapiere

ISSN 1617-5654

Weitere Publikationen des IAW:

- IAW-News (erscheinen 4x jährlich)
- IAW-Impulse
- IAW-Kurzberichte
- IAW Policy Reports
- IAW-Forschungsberichte

Möchten Sie regelmäßig eine unserer Publikationen erhalten, dann wenden Sie sich bitte an:

IAW Tübingen e.V.
Schaffhausenstraße 73, 72072 Tübingen
Telefon: 07071 98960
Fax: 07071 989699
iaw@iaw.edu

Aktuelle Informationen finden Sie auch im Internet unter:

http://www.iaw.edu

Wage expectation, information and the decision to become a nurse *

Philipp Kugler (IAW Tübingen)

11 January 2021

Abstract: In light of skilled-labor shortage in nursing, the effect of a change in the wage of nurses on their labor supply is intensely discussed in recent literature. However, most results show a wage elasticity close to zero. Using extensive data of former German 9th graders, I analyze the role of the expected wage as an incentive to become a nurse. To estimate a causal effect, I select controls and their functional form using *post-double-selection*, which is a data driven selection method based on regression shrinkage via the lasso. Contrary to common perceptions, the expected wage plays a positive and statistically significant role in the decision to become a nurse. Further, understating a nurse's wage decreases the probability of becoming one. Concerning omitted variable bias, I assess the sensitivity of the results using a novel approach. It evaluates the minimum strength that unobserved confounders would need to change the conclusion. The sensitivity analysis shows that potential unobserved confounders would have to be very strong to overrule the conclusions. The empirical results lead to two important policy implications. First, increasing the wage may help to overcome the shortage observed in many countries. Second, providing information on the (relative) wage may be a successful strategy to attract more individuals into this profession.

Keywords: health professional, expected wage, wage information, machine learning, sensitivity analysis
JEL-Classification: I11, I21, J24, J31
Correspondence: Philipp Kugler, Institute for Applied Economic Research, Schaffhausenstr. 73, 72072 Tübingen, Germany, philipp.kugler@iaw.edu

Declarations of interests: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*This paper is an outcome of the ZAFH care4care funded by the EFRE fund and the Ministry for Science, Research and Art Baden Wuerttemberg (grant number: FEIH_ZAFH_795924). I would like to thank Martin Biewen, Martin Kroczek and Jochen Späth for valuable discussions and comments. Yvette Bodry, Manuel Schick and Susanne Vögele are acknowledged for excellent research assistance. This paper uses data from the National Educational Panel Study (NEPS): Starting Cohort Grade 9, doi:10.5157/NEPS:SC4:10.0.0. From 2008 to 2013, NEPS data was collected as part of the Framework Program for the Promotion of Empirical Educational Research funded by the German Federal Ministry of Education and Research (BMBF). As of 2014, NEPS is carried out by the Leibniz Institute for Educational Trajectories (LIfBi) at the University of Bamberg in cooperation with a nationwide network.

1 Introduction

Due to demographic change and technological progress in medicine, the demand for skilled nurses has increased over the past decades (German Employment Agency 2020). This trend will continue in the coming years and will further aggravate the lack of nurses. To counteract this development, it is important to analyze and to understand the occupational behavior of nurses. The existing literature discusses a series of factors that might alleviate the lack of skilled workers. These include individual preferences of (future) nurses, improving working conditions and increasing wages. I contribute to this discussion by analyzing the effect of the ex-ante expected wage of young students on the probability of becoming a nurse. This is particularly interesting for at least two reasons: First, wages are the most controversially discussed factor in the literature. Some authors identify it as a very important factor influencing labor supply decisions of nurses (Hanel et al. 2014, Doiron et al. 2014). However, others suggest that the labor supply of nurses is relatively inelastic in terms of wages. Factors such as personal attitude and working conditions seem to play a much larger role (Shields 2004, McCabe et al. 2005). Second, although a central principle of the human capital theory suggests schooling decisions are made by comparing benefits and costs (Willis & Rosen 1979, Carneiro et al. 2011), research on the role of the expected wage as an incentive for choosing a nursing profession is scarce.

Since there are large differences in earnings depending on the occupational choices, the economic literature on the effect of the expected wage is rich (Altonji et al. 2016). The majority of studies agree that the wage has a significant and positive effect on the career choice (e.g. Boudarbat 2008, Montmarquette et al. 2002). Nonetheless, most studies find that preferences and interests play a larger role in career choice than the wage expectations (Beffy et al. 2012, Arcidiacono 2004).

In line with the economic literature, the nursing literature suggests preferences and interests to be the most important factors influencing the decision to become a nurse. In particular, caring for people is identified as the key reason for choosing the profession (e.g. Wilkes et al. 2015, Petrucci et al. 2016, Matthes 2019). Concerning the wage, several studies find that it only plays a minor role in the decision-making process (e.g. McCabe et al. 2005, Bomball et al. 2010, Cho et al. 2010). Based on these results, policy-makers might be tempted to focus on non-monetary factors to attract more young people into nursing. However, this contrasts recent work by Hanel et al. (2014) and Schweri & Hartog (2017). Schweri & Hartog (2017) examine the effect of ex-ante wage expectations on the decision to pursue a nursing degree (tertiary education) by using data on healthcare trainees (upper-secondary education) in Switzerland. Therefore, they analyze the decision on the intensive margin. Their results show that the greater ex-ante wage expectations of a nursing degree, the higher the probability to pursue such a degree later on. This indicates that higher wages may attract more students to become a high-skilled nurse. Hanel et al. (2014) estimate a model of labor supply decisions using data on individuals who hold a nursing qualification. The model accounts for the intensive and extensive margin by allowing individuals to enter and to exist occupations. As a result, they find a considerable high wage elasticity. This differs fundamentally from other work that detect very small elasticities (Shields 2004, Andreassen et al. 2017). These differences can be fully explained by the frequent neglect of the extensive margin and the exclusive analysis of the intensive margin. Although Hanel et al. (2014) do not account for the choice of becoming a nurse, their results suggest that wages may heavily drive the career choice, i.e. a decision on the extensive margin.

Using extensive panel data of former German 9th graders, this paper sheds further light on the role of wages in choosing a nursing profession. Evidence focusing on the influence of the wage on the probability to become a nurse is scarce and relies on the descriptive analysis of self-reported importance measures. So far, to my best knowledge, the effect of the (ex-ante) expected wage has only been considered by Schweri & Hartog (2017). However, they only analyze the impact of the expected wage among healthcare trainees (i.e. intensive margin). The data provides information on the wages that young students expect a nurse, a hairdresser, a motor vehicle mechanic, a bank clerk, a teacher and a physician to earn. Based on this information, I estimate the effect of the ex-ante expected wage on the probability to become a nurse among former 9th graders (i.e. extensive margin). Moreover, I estimate the effect of other factors (e.g. social orientation) on the probability of choosing the profession of a nurse. This allows to assess the magnitude of the impact of the expected wage and to fit my results into the recent literature. In addition, the data contains extensive background information on the individuals before their occupational decisions. This covers not only educational and parental background but also measures for personality, competencies, interests and attitudes. Overall, the data allows to observe over 150 characteristics. Hence, under the condition of unconfoundedness, I identify a causal effect of the expected wage on the probability to become a nurse. By applying the lasso proposed by Tibshirani (1996), a method that draws coefficients towards zero or exactly to zero, I am able to select the relevant controls and to model non-linearities in confounding. However, the lasso is tailored to choose variables such that an outcome is precisely predicted. Therefore, it cannot be applied directly for variable selection, when the aim is to identify a causal effect. As a solution, Belloni et al. (2012, 2014*b*) propose the *post-double-selection*, which is a two-step procedure to identify relevant controls and their functional form.

The central assumption to identify a causal effect is that no factors affecting the dependent variable and the variable of interest remain unobserved (unconfoundedness). This assumption is very strict. Although many potential confounders are available and included in the model, the assumption may not be credibly fulfilled. First, the occupational choice and formation of wage expectations are very complex processes. Hence, some unmeasured confounders may be left after including relevant controls. Second, by using *post-double-selection* I assume that only a small subset of all variables affect the career choice and the expected wage (approximate sparsity). If this assumption is violated, omitted variable bias might become an issue. Hence, I follow a novel approach by Cinelli & Hazlett (2020) to evaluate the sensitivity of the results regarding omitted variable bias. For linear models, they propose to assess the minimal strength that unobserved confounding needs to have on the expected wage and on the career choice in order to change the conclusion. To this end, Cinelli & Hazlett (2020) propose a procedure for benchmarking based on observed covariates. The knowledge about main predictors for career choice or the expected wage is the crucial premise for the benchmarking to be valuable. Fortunately, literature on determinants of wage expectations and factors driving young people into nursing is rich. Thus, credible benchmarking on observed covariates is possible.

This is by far not the only approach to assess the sensitivity of results. Several approaches exist. For example, in an influential paper, Oster (2019) proposes a method for computing the relative degree of selection on observed and unobserved variables to match a given treatment effect (which is zero, for example). However, the degree of relative selection is hard to grasp and interpret. Moreover, the computation requires the specification of the unknown maximum explanatory power that can be achieved by a regression of the outcome on both observed and unobserved controls. By contrast, the method by Cinelli & Hazlett (2020) only relies on quantities that are easy to understand and interpret.

My results show that the expected wage plays a positive, statistically significant role in the decision to become a nurse. In line with recent literature, individual preferences play a larger role than the expected wage. Since the career choice is a decision on the extensive margin, my results are also consistent with those of Hanel et al. (2014). The importance of the extensive margin is further underlined by the result that effects are driven by young people who do not become a nurse and underestimate the wage. This means that the public perception of wages in nursing is too low. Therefore, nursing is less attractive than other occupations for which wages are not systematically understated. To combat the lack of skilled nurses, policy-makers can make the profession more attractive by increasing the relative expected wage of a nurse.

The remaining paper is structured as follows. Section 2 outlines the methods applied in the empirical analysis. Section 3 briefly describes the data, the expected wage measures as well as the control variables. In section 4, I present and discuss the main results of my analysis. Section 5 concludes.

2 Methods

2.1 Post-Double-Selection

The causal effect of the expected wage w_i on the probability to become a nurse is estimated by a partially linear model

$$y_i = \beta w_i + g(x_i) + \zeta_i, \tag{1}$$

where $y_i \in \{0, 1\}$ denotes the binary choice to become a nurse. The function $g(x_i)$ is unknown and potentially complicated. I approximate it by a linear combination that may include higher order polynomials and interactions

$$g(x_i) = x_i'\theta_y + r_{yi}, \tag{2}$$

where r_{yi} is an approximation error. The aim is to estimate β. In order to conduct inference about it and interpret it as a causal effect, I need to rely on the assumption of unconfoundedness $\mathbb{E}\left[\zeta_i | w_i, r_{yi}, x_i\right] = 0$. It states that all factors that affect the choice y_i and the expected wage w_i at the same time must be contained in $g(x_i)$. However, it is a difficult task to define a set of variables to be included in the model and to model their functional form (i.e. what polynomials and interactions to include). Therefore, I rely on data-driven variable selection and follow the *post-double-selection* (PDS) approach proposed by Belloni et al. (2012, 2014*b*). The lasso is a shrinkage method that imposes a penalty on the size of the coefficients, i.e. shrinks them towards zero or exactly to zero. This prevents models with many variables that are correlated with each other from overfitting (Hastie et al. 2009). The lasso is defined as

$$\hat{\gamma}^{lasso} = \arg\min_{\gamma} \Bigg\{ \underbrace{\frac{1}{2} \sum_{i=1}^{N} \Bigg(y_i - \gamma_0 - \sum_{j=1}^{p} x_{ij}\gamma_j \Bigg)^2}_{\text{residual sum of squares}} + \underbrace{\lambda \sum_{j=1}^{p} |\gamma_j|}_{\text{penalty term}} \Bigg\}, \tag{3}$$

where $\sum_{j=1}^{p} |\gamma_j|$ imposes the penalty on the size of the coefficients and the parameter $\lambda \geq 0$ controls the magnitude of the punishment.

A naive approach to estimate β would be to apply the lasso estimator to equation (1) and to exclude β from the penalty term such that it is enforced to stay in the model. Afterwards one might use a least-squares regression of the outcome on w_i and controls with non-zero coefficients. However, this approach leads to biased estimates because of omitted variables. The lasso is designed to learn a forecasting rule of y_i given w_i and x_i and not to learn about the causal relationship between y_i and w_i given controls x_i (Belloni et al. 2014*a*). Therefore, lasso cannot be used off the shelf for the estimation of causal effects. As a solution, Belloni et al. (2012, 2014*b*) propose an intuitive and easy-to-implement procedure. First, the lasso is used to estimate a model predicting the outcome given x_i in equation (4) and a further model predicting the expected wage given x_i in equation (5)

$$y_i = x_i'\pi + \epsilon_i, \tag{4}$$

$$w_i = x_i'\theta_w + \nu_i. \tag{5}$$

Subsequently, all variables with non-zero coefficients in either of the two models are kept as control variables in order to estimate $\hat{\beta}$ in equation (1) by an ordinary least squares regression. This step is known as the "post-lasso". The crucial assumption under which PDS works is approximate sparsity. It states that the expected wage and the career choice can be approximated by equation (4) and (5) using only a small number of covariates relative to the sample size. Additional variables that are considered as important for ensuring robustness, can be included (amelioration set). The condition is that the amelioration set is not substantially larger than the number of variables chosen via the lasso (Belloni et al. 2014*b*).

The choice of λ is of importance. With the aim of prediction, standard lasso applications choose λ by cross-validation. However, this analysis aims to describe the relationship between career choice and the expected wage. If λ is too large, only a few variables are selected and omitted variable bias may occur. If λ is too small, the number of variables is very large such that overfitting may become an issue. Therefore, I follow Urminsky et al. (2016) and use $\lambda = 1.1\sigma_R \frac{1}{\sqrt{N}} \Phi^{-1}(1 - \frac{0.1}{ln(N)2p})$, where N is the number of observations, p is the number of potential controls, Φ^{-1} denotes the inverse cumulative function of the standard normal distribution and σ_R the standard deviation of the residuals of the model. Finally, it is important to note that the chosen variables are not interpretable since selection depends on the sample (Mullainathan & Spiess 2017).

2.2 Sensitivity

Even though I have access to an extensive set of potential controls x_i, bias due to unobserved confounders cannot be ruled out. For example, covariates measuring the interests of the individuals might not fully capture all relevant aspects but only a share of it. Moreover, the assumption of approximate sparsity may be violated. There may exist covariates that are not selected by lasso but affect both, the expected wage and the decision to become a nurse. Therefore, I make use of a procedure proposed by Cinelli & Hazlett (2020) to analyze the sensitivity of the results due to potentially unobserved (non-)linear confounding factors z. In a nutshell, they propose to assess the sensitivity of the estimates by analyzing whether a confounder is strong enough to change the conclusion if it is as strong as a very good predictor of y or w.

Conventionally, the omitted variable bias can be written as $\widehat{bias} = \hat{\gamma}\hat{\delta}$. Hence, $\hat{\gamma}$ describes the difference in the linear expectation of the outcome if z_i changes by one unit, holding everything else constant and $\hat{\delta}$ describes the difference in linear expectation of the confounder if the variable of interest changes by one unit, holding everything else constant (Cinelli & Hazlett 2020). Arguing that both quantities $\hat{\delta}$ and $\hat{\gamma}$ are hard to grasp, Cinelli & Hazlett (2020) write the conventional omitted variable bias formula in terms of partial R^2 measures. Those are easier to interpret and can be exploited for further analysis. Denote $\hat{\beta}_{obs}$ as the observed estimated effect and $\hat{\beta}$ as the estimated effect from a model controlling unobserved confounding factors, i.e. $\hat{\beta} = \hat{\beta}_{obs} - \widehat{bias}$. Then, they show that

$$|bias| = \hat{se}(\hat{\beta}_{obs})\sqrt{\frac{R^2_{y\sim z|w,x} R^2_{w\sim z|x}}{1 - R^2_{w\sim z|x}} df}, \tag{6}$$

where df defines the degrees of freedom, $R^2_{y\sim z|w,x}$ stands for the partial R^2 of regressing y on z after controlling for w and x and $R^2_{w\sim z|x}$ denotes the partial R^2 of regressing w on z after controlling for x. Further, the standard error of $\hat{\beta}$ can be written as

$$\hat{se} = \hat{se}\left(\beta_{obs}\right)\sqrt{\frac{1-R^2_{y\sim z|w,x}}{1-R_{w\sim z|x}}\left(\frac{df}{df-1}\right)}, \tag{7}$$

and the adjusted t-statistic is defined as $t_{adj} = \hat{\beta}/\hat{se}$. Applying these definitions, $\hat{\beta}, \hat{se}$ and t_{adj} can be computed by substituting reasonable values for $R^2_{y\sim z|w,x}$ and $R^2_{w\sim z|x}$, i.e. the strength of confounding, into equations (6) and (7). However, actual knowledge about the absolute strength is seldom available. As a solution, Cinelli & Hazlett (2020) argue that the researcher is often able to make a statement on the relative strength of potential unobserved confounding, e.g. z cannot account for as much variation of the outcome as some observed covariate x_j. There are several ways to formalize such claims. I follow Cinelli & Hazlett (2020) and claim that I measure the key determinant of y and w such that the omitted variable cannot explain as much residual variance in y or w as this determinant. Define

$$k_w = \frac{R^2_{w\sim z|x_{-j}}}{R^2_{w\sim x_j|x_{-j}}} \tag{8}$$

$$k_y = \frac{R^2_{y\sim z|x_{-j},w}}{R^2_{w\sim x_j|x_{-j},w}}, \tag{9}$$

where x_{-j} is a vector including all variables contained in x, excluding x_j. The ratios k_w and k_y show how much of the variance in w or y is explained by z relative to the explanatory power of x_j, conditional on all other covariates. In this paper $k_w = k_y = 1$, i.e. I consider the impact of a confounder z that is as strong as x_j. Given k_w and k_y, Cinelli & Hazlett (2020) show that

$$R^2_{w\sim z|x} = k_w f^2_{w\sim x_j|x_{-j}} \quad R^2_{y\sim z|w,x} \leq k_y \eta^2 f^2_{y\sim x_j|x_{-j},w}, \tag{10}$$

where η is a scalar that depends on k_w, k_y, and $R^2_{w\sim x_j|x_{-j}}$. Furthermore, $f^2_{w\sim x_j|x_{-j}}$ denotes partial Cohen's f of w on x_j and $f^2_{y\sim x_j|x_{-j},w}$ denotes partial Cohen's f of y on x_j.[1] Cinelli & Hazlett (2020) have shown that these robustness results are exact for a single linear confounder and conservative for multiple, possibly nonlinear, confounding factors.

It is important to emphasize that this bounding procedure heavily relies on the choice of the benchmark variable x_j. If it is not true that x_j is a key predictor of the outcome or treatment, the bounding is pointless. Hence, domain knowledge is necessary (Cinelli & Hazlett 2020). In the following, I choose observed covariates that are often discussed in the literature. First, bounding is

[1] Note that Cohen's f^2 is defined as $f^2 = \frac{R^2}{1-R^2}$.

based on social orientation. It is the key characteristic of those who become a nurse (e.g. Matthes 2019), while preferences are generally a decisive factor in career choice (e.g. Arcidiacono 2004). In addition, interests also play an important role in the formation of expected wages (Wiswall & Zafar 2015). Second, the professions of the parents play an important role in the occupational choice (e.g. Knoll et al. 2017). Therefore, the results are bounded by an indicator that indicates whether at least one of the parents is a nurse. Moreover, parents in nursing might inform their children about the expected wages. Third, an indicator for gender is considered. Females become nurses much more often than males (Speer 2019). Moreover, gender also plays a crucial role in wage expectations: females expect lower wages than males (e.g. Brunello et al. 2004, Fernandes et al. 2020). Fourth, (perceived) ability determines the expected wages (Brunello et al. 2004). Therefore, a measure for ability, namely metacognition, is used to bound the results. Note, that these variables have to be part of the model in order to use them as benchmark variables. Hence, the amelioration set contains these four variables, to ensure that they are not excluded by data-driven variable selection.

3 Data

This study uses Starting Cohort Four (SC4) of the German National Educational Panel Study (NEPS). The survey collects data on young people who attended the 9th grade in 2010 and has been followed since (Blossfeld & von Maurice 2011). For several reasons the data is highly suitable for investigating the role that the expected wage plays in the decision to become a nurse. Since the data is available from 2010 to 2016, the transition from school to further education can be observed in great detail and no retrospective information has to be used. The following analysis focuses on the choice of the first occupational training, which certainly has a relevant impact on the further life course. Beyond that, the individuals are asked how much they expect to earn as a nurse, a hairdresser, a motor vehicle mechanic, a bank clerk, a teacher and a physician.[2] In order to define a measure for the expected wage of nurses, the expected wages of all six occupations are ranked from lowest to highest. If the wage cannot be assigned unambiguously due to ties, the mean rank is assigned such that the sum of ranks is preserved. Formally, I define the expected

[2]The following question is asked: "Now, we are also interested in your estimate of the amount of wages paid in certain jobs. What is in your opinion the monthly pay of ...?"

wage rank of a nurse as

$$\text{rank}_i^{\text{nurse}} = 1 + \sum_{w_i \in \{S_i \setminus w_i^{\text{nurse}}\}} \mathbb{1}\left(w_i < w_i^{\text{nurse}}\right) + 0.5 \times \mathbb{1}\left(w_i = w_i^{\text{nurse}}\right), \tag{11}$$

where $\mathbb{1}(\cdot)$ denotes an indicator function that takes the value 1 if the expression in the parentheses is true, S_i is the set of surveyed expected wages and w_i^{nurse} is the expected wage of a nurse. Two further measures of the relative expected wage of a nurse are defined as the ratio between individual's i expected wage as a nurse and maximum as well as minimum specified wage

$$\text{relwage}_i^{\text{nurse, max}} = \frac{w_i^{\text{nurse}}}{w_i^{\text{max}}}, \tag{12}$$

$$\text{relwage}_i^{\text{nurse, min}} = \frac{w_i^{\text{nurse}}}{w_i^{\text{min}}}. \tag{13}$$

In addition, I use the expected absolute wage of a nurse.

Based on the ranking measure in equation (11), I can easily assess how close the relative wage expectations are to reality by computing the deviation from the true ranks. The median wages reported by German Employment Agency (2018) provide the basis for the true rank. According to this source of information, the following *true* ranking from lowest to the highest wage was established: (1) hairdresser, (2) motor vehicle mechanic, (3) nurse, (4) bank clerk, (5) teacher and (6) physician. The ranking is utilized to construct a measure that captures the knowledge about relative wages by adding the absolute deviations of the expected rank of each occupation

$$\begin{aligned} \text{rank}_i^{\text{abs. dev.}} =& |\text{rank}_i^{\text{barber}} - 1| + |\text{rank}_i^{\text{mechanic}} - 2| + |\text{rank}_i^{\text{nurse}} - 3| + \\ & |\text{rank}_i^{\text{bank clerk}} - 4| + |\text{rank}_i^{\text{teacher}} - 5| + |\text{rank}_i^{\text{physician}} - 6|, \end{aligned} \tag{14}$$

where the ranks of each occupation are computed the same way as the rank of a nurse's wage. Additionally, I can construct indicators that show whether the wage rank of a nurse is overestimated, correctly estimated or underestimated

$$\text{rank}_i^{\text{nurse, over}} = \mathbb{1}\left(\text{rank}_i^{\text{nurse}} > 3\right), \tag{15}$$

$$\text{rank}_i^{\text{nurse, correct}} = \mathbb{1}\left(\text{rank}_i^{\text{nurse}} = 3\right), \tag{16}$$

$$\text{rank}_i^{\text{nurse, under}} = \mathbb{1}\left(\text{rank}_i^{\text{nurse}} < 3\right). \tag{17}$$

Besides information on expected wages, there are other potentially important factors available that may drive young people into or out of nursing (see Wohlkinger et al. 2011). This enables me to assess the importance of the expected wage by comparing the effect with other effects estimated in the literature. A large share of recent work finds that those who become nurses do not rate the importance of economic factors as important as those who choose another profession. Therefore, I use a measure of the importance of economic factors (i.e. risk of unemployment and financial aspects) in choosing a career. Moreover, helping others is considered to be one of the main driving forces in choosing a nursing profession. Hence, a measure that quantifies the amount of social interests is used. Finally, I estimate a model that uses an indicator of self-assessed importance of comfort (i.e. physical working conditions).

Additionally, extensive information about the background, personal characteristics and the (social) environment of the individuals are measured before they have decided on a career. Such extensive information is crucial for the identification of causal effects. All potential controls are summarized in table B1. The exclusion of observations with at least one missing value would lead to a substantial loss of observations. Therefore, I impute missing values by chained equations (van Buuren & Groothuis-Oudshoorn 2011). To estimate the impact of the expected wage via equation (1), I account for non-linearities in confounding by interacting all variables with each other and by additionally including fifth-order polynomials of non-binary covariates. As a result, 13.878 potential controls are available.

After excluding individuals with extreme wage expectations, missing values in variables of interests[3] or with too many missing observations in general[4], I observe 7089 individuals that transition from school to occupational training, of whom 238 chose nursing.

4 Results

4.1 What does the expected wage capture?

Before the results are presented, it is necessary to clarify exactly what the expected wage measures capture. Individuals are asked to state their *opinion on the monthly salary of a nurse*. Consequently, I measure the information about nurses' wages. This may include knowledge

[3]That is expected wage, economic and social orientation and importance of comfort.

[4]Precisely, I drop observations with over 18 % missing values - that is the 90 % quantile of the share of missing entries

about average wages, knowledge of wages according to collective agreements, but also wrong expectations due to the lack of information or wrong perceptions of wages. This differs from a large share of the literature on wage expectations that often uses information on realized future earnings (e.g. Montmarquette et al. 2002, Beffy et al. 2012) or directly asks young people what they expect to earn after graduation (e.g. Brunello et al. 2004). The question at hand does not capture beliefs about future wages based on knowledge and perceptions about individual skills. However, knowledge and perception of skills and knowledge about wages may easily get mixed up. For example, individuals who possess high social skills and therefore want to become a nurse may inform themselves about wages in leadership positions as they expect to rise fast and earn more than the average. Further reasons may exist. I address this by having access to an extensive set of potential controls that measure skills, interests and several other background characteristics. Therefore, the estimated effects only contain information about wages.

4.2 The role of the expected wage

4.2.1 Descriptive evidence

First, I provide some insights about the univariate relationship between wage expectations and the choice whether or not to become a nurse. Table 1 depicts the shares of the nurse's expected wage ranks reported by nurses and others. The ranks of both groups seem to follow a similar general

Table 1 – Distribution of the expected wage rank

	Expected nurse's wage rank					
	1	2	3	4	5	6
nurses	5.88	39.07	36.13	13.87	4.62	0.42
others	14.86	48.65	27.63	6.38	1.96	0.53
all	14,55	48,32	27,92	6.63	2.05	0.52

The table depicts the share of the expected wage rank by nurses and others. For the sake of clarity, in the case of ties, the lower rank is reported.

pattern. Both most often expect nursing to be the second and the third rank. Further, both rarely expect nurses to have the lowest earnings or expect a rank higher than three. However, the specific patterns strongly differ. The respective shares of future nurses who expect a rank larger than two exceed the shares of the others. Furthermore, the shares for the first and second rank are smaller among nurses. To gain further insights into the differences in expected wage ranks, table 2 presents the mean differences of the expected wage rank by nurses and others. Table 2 reveals that, on average, nurses expect a lower rank of the wage earned by hairdressers or mechanics than others, but expect a larger rank of a nurse's wage. Interestingly, the average expected ranks of a bank clerk, teacher and physician are not significantly different between nurses and others. This means, that expectations only differ for occupations with lower wages.

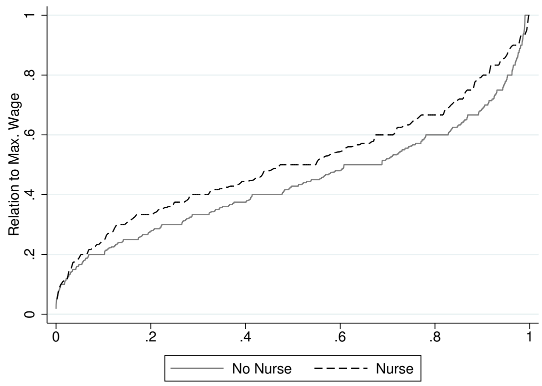
Table 2 – Differences in expected mean wage ranks by nurses and others

	Mean of others	Mean of nurses	Difference	P-values of test for differences in means		
Expected rank of a...				H_0 : diff. < 0	H_0 : diff. $= 0$	H_0 : diff. > 0
... barber	1.29	1.16	0.12	1.00	0.00	0.00
... mechanic	2.74	2.56	0.18	1.00	0.00	0.00
... nurse	2.46	2.84	-0.39	0.00	0.00	1.00
... bank clerk	4.53	4.49	0.04	0.73	0.54	0.27
... teacher	4.41	4.35	0.06	0.83	0.33	0.17
... physician	5.56	5.58	-0.14	0.40	0.80	0.60

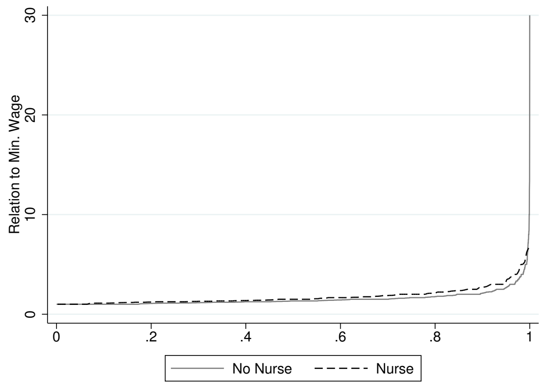
The table depicts the means of expected wage ranks by future nurses and others together with their differences. Further, to assess if differences are statistically significant, t-tests are conducted.

In figures 1a-1c the remaining expected wage measures are depicted in quantile plots. Figure 1a shows the ratio of the expected wage of a nurse and the highest expected wage as defined in equation (12). Differences in relative expected wages between those who become a nurse and those who do not, are very clear. Except for the lower and upper end, future nurses expect higher relative wages. Similarities in lower and upper ends indicate that extreme expectations do not differ systematically between groups. In figure 1b, the distribution of the ratio between a nurse's expected wage and the lowest expected wage is shown. For non-future nurses the extreme value at the upper end of the distribution is prominent. However, this appears to be an outlier. In the remaining distribution, the relative expected wage is larger for future nurses. Differences increase in higher quantiles. In summary, descriptive evidence consistently suggests that future nurses expect a higher relative wage than those who do not become a nurse. As figure 1c reveals, not only the relative expected wage of a nurse is higher for future nurses. At least in quantiles in the middle, the absolute expected wage is also slightly higher.

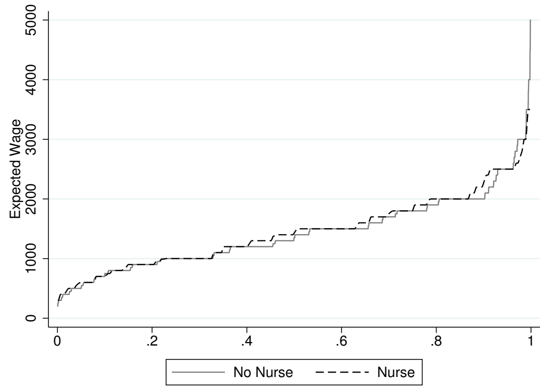
Figure 1 – Continuous expected wage measures



(a) Ratio of expected wage of a nurse and maximum expected wage



(b) Ratio of expected wage of a nurse and minimum expected wage



(c) Expected absolute wage of a nurse

Each panel depicts a quantile plot of one wage measure. The ratio of the expected wage of a nurse and maximum expected wage is defined as $\text{relwage}_i^{\text{nurse, max}} = w_i^{\text{nurse}}/w_i^{\text{max}}$ and ratio of the expected wage of a nurse and minimum expected wage is defined as $\text{relwage}_i^{\text{nurse, min}} = w_i^{\text{nurse}}/w_i^{\text{min}}$

4.2.2 Results of PDS

The observed descriptive differences may be caused by confounding. For example, those who have no interests in becoming a nurse may expect a rather low wage (e.g. Wiswall & Zafar 2015). However, the aim of the analysis is to uncover whether the expected wage plays a role in becoming a nurse, given the characteristics of the individuals (i.e. equally interested, same background, same skills, etc.). As described in section 2, I tackle this issue by using PDS to estimate the effect of the wage expectation on the probability to become a nurse. The results are depicted in table 3. Each of the three columns shows the results of an unconditional model, i.e. single OLS, and the post-lasso, i.e. a conditional OLS model with controls chosen by double-selection. In the first column, future nurses are compared to all remaining young people. However, this neglects the heterogeneity of the impact of expected (relative) wages. Individuals who are interested in becoming a nurse, e.g. chose a similar occupation, may be more responsive to expected wages compared to those who have no interest in nursing at all. Hence, in the remaining columns the sample is restricted regarding the career choices. I compare future nurses to (2) young people who opted for vocational training and (3) individuals who chose a social field.[5] Each panel of the table depicts the results of one of the four expected wage measures described above. The measures are standardized to have mean 0 and standard deviation 1 such that the results can easily be compared with other factors in the subsequent section.[6]

The first panel shows the estimated effects of a nurse's rank on the probability to become a nurse. Column one compares future nurses to all the remaining individuals in the data. As expected from the descriptive results in tables 1 and 2, results of the unconditional model show that an increase of the rank by one standard deviation is associated with a statistically significant increase in the probability to become a nurse by 1.4 percentage points. When relevant controls are added, I still observe a statistically significant change by 1 percentage point. The slightly smaller coefficients of the post-lasso compared to the single OLS model shows that those, who are prone to become a nurse (e.g. having parents that are nurses and young people that have a social attitude) expect higher relative wages. The results change only slightly with regard to the comparison group in column two. Comparing future-nurses with individuals who chose a social field, I find a much larger effect in both the unconditional and conditional model. These results hint to heterogeneity in the effect, i.e. larger effects for those who chose a more similar field. Interestingly, the number of included controls is much smaller when the comparison group only consists of individuals who chose a vocational training or a social field. However, this is expected since the sample size is much smaller. Thus, λ becomes larger and draws the coefficients of the lasso models stronger towards zero. Further, the sample in column 1 is more heterogeneous than the ones in column 2 and 3. Therefore, fewer variables may be required to explain differences.[7]

[5]Note, comparing future-nurses with youths who chose a vocational training, i.e. did not choose to visit a university, is motivated by the German education system. Education after school is divided into academic and vocational training, whereas nursing belongs to the latter kind.

[6]A discussion of the magnitude of the estimates will be given in the next section.

[7]Note that regarding the choice of variables, mainly interactions are chosen. This hints to strong non-linearities in confounding and stresses the importance of flexible choice of controls.

Table 3 – Expected wage of a nurse

	(1)		(2)		(3)	
	nurse vs. all		nurse vs. vocational training		nurse vs. social field	
	Single OLS	Post-Lasso	Single OLS	Post-Lasso	Single OLS	Post-Lasso
Rank of nurse's wage						
	0.014***	0.010***	0.015***	0.012***	0.047***	0.035***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.008)	(0.009)
p	-	41	-	25	-	13
Nurse's wage/highest wage						
	0.011***	0.006***	0.009***	0.006*	0.036***	0.023***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.009)	(0.009)
p	-	38	-	16	-	20
Nurse's wage/lowest wage						
	0.009***	0.008***	0.011***	0.011***	0.025***	0.022***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.008)	(0.007)
p	-	24	-	19	-	9
Nurse's absolute wage						
	0.002	0.008***	0.006**	0.011***	0.021**	0.028***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.009)	(0.009)
p	-	29	-	23	-	13
N	7098		4452		1616	

The table depicts the results of the effect of the expected wage on the decision to become a nurse. The expected rank of nurse's wage is defined in equation (11) and the ratio of expected wage as a nurse and highest/lowest expected wage is defined in equation (12) and (13) respectively. Standard errors are shown in parentheses. Significance of the coefficient at conventional significance levels $1\%, 5\%, 10\%$ are indicated by stars ***, **, * respectively. N indicates the number of observations and p the number of chosen controls.

The second panel contains the results of the effect of the ratio of an expected nurse's wage and the highest expected wage. Results for the entire sample in the first column show that even after controlling for relevant confounders, I find a statistically significant and positive effect on the probability to become a nurse. Similar to results of the wage rank, the coefficient in the post-lasso model is smaller than in the unconditional model. The effect stays positive and significantly different from zero when the comparison group is changed. For those who chose a social field in column 3, the effect is again much larger. This suggests effect-heterogeneity.

The following panel shows the impact of the ratio of the expected nurse's wage and the lowest expected wage on the decision to become a nurse. Comparing those who become nurses with all other individuals, results of the unconditional model show that an increase in the relative wage increases the probability to become a nurse by 0.9 percentage points. When relevant controls are included, the probability increases by 0.8 percentage points. Just like the estimates in the first and second panel, the results also indicate heterogeneity. The effects become even larger when the comparison group consists of young people who chose a social field. Figure 1b revealed that some non-future nurses expect an outlying high ratio between a nurse's wage and the lowest wage. These outliers do not qualitatively alter the results. Their exclusion, if anything were to change, would cause even larger effects.

The last panel of table 3 contains the results of the impact of the expected absolute wage on the probability to become a nurse. After conditioning on relevant controls, I find a statistically significant effect of the absolute wage on the probability to become a nurse that stays significant when the composition of the comparison group is changed. Interestingly, in all columns, the coefficient in the model with no controls is smaller than in the models including controls. Whereas those who are prone to become a nurse expect higher relative wages, they expect a lower absolute wage. As observed for the relative wage measures, the effect becomes larger when the comparison group is restricted to individuals who chose a similar occupation.

In summary, results in table 3 show that even after conditioning on an extensive set of relevant controls and accounting for non-linearities in confounding, the expected nurse's wage affects the probability to become a nurse. This holds true for both the relative and absolute wage. Moreover, I find evidence that effects are heterogeneous. These effects are stronger for individuals who are more prone to choose a social occupation.

4.2.3 Sensitivity of the post-lasso

As outlined previously, the presence of omitted variable bias may be likely if the assumption of approximate sparsity is violated or if unobserved confounders remain despite a large set of controls. To analyze the impact of potential unobserved confounding, I conduct a sensitivity analysis as described in section 2. The results are shown in table 4. Each panel depicts the results of one wage measure. As discussed above, I follow Cinelli & Hazlett (2020) and make use of observed covariates that are strong predictors of the occupational choice or the expected wage to analyze the consequence of potential omitted variables. Columns 2 to 5 display the adjusted estimate $\hat{\beta}$ and t-statistic t_{adj}. They are obtained when an unobserved confounder, that explains as much variance in y and w as predefined benchmark variables, is additionally controlled for. As mentioned above, the variables used to bound the consequences of omitted variable bias are (1) gender, (2) parents' occupation, (3) social interests and (4) metacognition. The first column indicates whether $\hat{\beta}$ and t_{adj} are computed using only one variable or whether it is based on all transformations in the model involving the variable. For example, the adjusted estimate and t-statistic with no transformations are computed under the assumption that an unobserved confounder that is as strong as gender exists. In contrast, $\hat{\beta}$ and t_{adj} including transformations are computed by assuming that an unobserved confounder exists, that is as strong as gender and all interactions that are included in the model and where gender is involved in (e.g. interaction between gender and math-skills, gender and social interests, etc.). I expect benchmarks that account for transformations to have a much larger impact than benchmarks of single covariates, because many transformations are chosen by the lasso. The last column shows the results that would have been obtained if an omitted variable that explains as much as all four variables together had been controlled for.

Table 4 – Sensitivity analysis

	(1)	(2)		(3)		(4)		(5)		(6)	
	Including transformations	Gender: female		Parents nurses		Social interests		Metacognition		All	
		$\hat{\beta}$	t_{adj}	$\hat{\beta}$	t_{adj}	$\hat{\beta}$	t_{adj}	$\hat{\beta}$	t_{adj}	$\hat{\beta}$	t_{adj}
Rank of nurse's wage											
nurse vs. all	No	0.009	4.43	0.009	4.39	0.009	4.38	0.010	4.45	0.009	4.28
	Yes	0.008	3.92	0.008	3.87	0.008	3.97	0.009	4.40	0.005	2.38
nurse vs. vocational training	No	0.012	3.75	0.012	3.78	0.012	3.75	0.012	3.79	0.011	3.62
	Yes	0.010	3.19	0.010	3.19	0.010	3.26	0.011	3.57	0.004	1.20
nurse vs. social field	No	0.035	3.94	0.035	3.97	0.035	3.97	0.035	3.97	0.034	3.85
	Yes	0.035	3.94	0.031	3.59	0.035	3.97	0.035	3.96	0.030	3.43
Nurse's wage/highest wage											
nurse vs. all	No	0.006	2.77	0.006	2.79	0.006	2.74	0.006	2.79	0.006	2.62
	Yes	0.005	2.44	0.005	2.30	0.005	2.29	0.006	2.79	0.002	1.12
nurse vs. vocational training	No	0.006	1.85	0.006	1.90	0.006	1.87	0.006	1.90	0.005	1.70
	Yes	0.005	1.59	0.005	1.51	0.004	1.40	0.006	1.88	0.001	0.30
nurse vs. social field	No	0.023	2.56	0.023	2.57	0.022	2.44	0.023	2.58	0.021	2.40
	Yes	0.023	2.56	0.020	2.27	0.022	2.44	0.023	2.58	0.018	2.06
Nurse's wage/lowest wage											
nurse vs. all	No	0.008	3.57	0.007	3.48	0.008	3.57	0.008	3.58	0.007	3.43
	Yes	0.007	3.50	0.007	3.37	0.007	3.15	0.007	3.54	0.005	2.62
nurse vs. vocational training	No	0.010	3.26	0.010	3.18	0.010	3.27	0.010	3.27	0.010	3.10
	Yes	0.010	3.22	0.010	3.04	0.009	2.84	0.010	3.23	0.008	2.46
nurse vs. social field	No	0.021	2.90	0.022	3.01	0.020	2.77	0.022	2.99	0.019	2.61
	Yes	0.021	2.90	0.021	2.95	0.020	2.77	0.022	2.99	0.017	2.38
Nurse's wage											
nurse vs. all	No	0.008	3.52	0.007	3.49	0.008	3.51	0.008	3.52	0.007	3.43
	Yes	0.006	2.67	0.007	3.21	0.007	3.18	0.007	3.52	0.003	1.20
nurse vs. vocational training	No	0.011	3.43	0.011	3.46	0.011	3.42	0.011	3.43	0.011	3.26
	Yes	0.008	2.54	0.010	3.18	0.010	3.10	0.011	3.42	0.003	0.81
nurse vs. social field	No	0.028	3.12	0.027	3.07	0.028	3.12	0.028	3.12	0.026	2.96
	Yes	0.025	2.84	0.025	2.85	0.028	3.12	0.028	3.12	0.019	2.14

The table depicts the results on the sensitivity of the effect of the expected wage on the decision to become a nurse. The adjusted t-statistic is based on the adjusted estimate $\hat{\beta}$ and adjusted standard errors $\hat{se}$. $R^2_{w\sim z|x}$ and $R^2_{y\sim z|w,x}$ are computed as defined in equation (10) setting $k_w = k_y = 1$, i.e. unobserved confounders that are as strong as the considered benchmark variables.

The first panel depicts the sensitivity of the results regarding the expected rank of a nurse's wage. The adjusted estimate only decreases slightly and equals 0.009, provided that there exists an unobserved confounder as strong as gender for which is additionally controlled. The change in the adjusted t-statistic is very small such that results stay significant at a 1% significance level. Confounders as strong as parent's occupation, social interests and metacognition only lead to minor changes. Even if I additionally control for a confounder that is as strong as all four benchmark variables combined, the conclusion that the expected wage rank of a nurse significantly affects the choice to become a nurse is still valid. As expected, adjusted estimates $\hat{\beta}$ are drawn to zero by a larger amount when transformations are included. Nonetheless, these changes are small. The effect decreases to 0.004 if I control for a confounder that is as strong as all four benchmark variables together and includes all their transformations. It remains significant at the 5%-level. Therefore, the estimated effect is not sensitive to potential unobserved confounding. A change in the comparison group leads to similar robust results. The only noteworthy change in the conclusion occurs when the comparison group consists only of those who chose a vocational training. It is caused by a confounder that is as strong as all four benchmark variables including their transformations. The adjusted t-statistic shows that if such a confounder exists, there is no statistically significant effect anymore.

The second panel displays the sensitivity of the results on the ratio between the expected nurse's wage and the highest wage. The results for the entire sample show that only controlling for a confounder that is as strong as all four benchmark variables and their respective transformations has an impact that is large enough to change the conclusion. The effect decreases to 0.002 and is not statistically significant. The sensitivity analysis reveals that the estimated effect is sensitive when the comparison group consists of individuals undergoing a vocational training. A confounder as strong as single variables is not strong enough to change the conclusion. However, a cofounder as strong as gender, parental occupation or interests together with their respective transformations leads to an effect that is not statistically significant different from zero. It is evident that a confounder, as strong as all four benchmark variables combined and including their transformation, leads to an insignificant effect too. The result of comparisons between nurses and individuals in a social field are not sensitive to any of the considered strengths of confounding.

The third panel depicts the sensitivity of the ratio between a nurse's wage and the lowest wage. The results show that no confounder as strong as the considered benchmark variables is strong enough to change the conclusion. Even a confounder as strong as all four benchmark variables together including their transformation does not lead to remarkable changes in the estimated effect. Similar sensitivity results can also be observed when the comparison group is changed.

The last panel shows how a confounder changes the estimated effect of the expected nurse's wage. A confounder as strong as a single variable does not have an impact on the estimated effect. Even a confounder as strong as all four benchmark variables together does not change the estimated effect. However, the impact of a confounder as strong as all four benchmark variables including their transformations is considerable. The effect decreases to 0.003 and is not significantly different from zero. The results are similar when the comparison group is changed. Comparing nurses to individuals who chose a vocational training, the impact of a confounder as strong as all four variables combined including their transformations is strong enough to change the conclusion. The effect substantially decreases to 0.003 and is not significantly different from zero. Choosing individuals in a social field as comparison group, none of the considered confounders is strong enough to change the conclusion.

Taking into account that the gender, the parents and the interests are key drivers of occupational choice, it can be concluded that results of the expected wage rank, the ratio between a nurse's wage and the lowest wage and the absolute wage are only sensitive to a confounder that is very strong. Similarly, the results on the effect of the ratio between the nurse's wage and the highest wage are only sensitive regarding a very strong confounder when comparing nurses to all other individuals or to those who chose a social field. However, the results are sensitive when nurses are compared to those who chose a vocational training. If a confounder with a certain strength exists, only subgroups are affected by the ratio.

4.3 How much do other factors matter?

In order to assess the size of the effects of the expected wage and to obtain some reassurance about the validity of the data, I compare the effect to other factors discussed in the recent literature. More precisely, I estimate three further PDS models using the self-assessed importance of economic factors, social interests and self-assessed importance of comfort aspects instead of the expected wage. The results are depicted in table 5. To compare the size of the effects with the effect of the expected wage in table 3, measures are standardized to have mean 0 and standard deviation 1.

Table 5 – Relevance of other factors

	nurse vs. all		nurse vs. vocational training		nurse vs. social field	
	Single OLS	Post-Lasso	Single OLS	Post-Lasso	Single OLS	Post-Lasso
Importance of economic factors						
	-0.002	0.000	-0.006*	0.000	0.008	0.015
	(0.002)	(0.003)	(0.003)	(0.004)	(0.008)	(0.010)
p	–	52	–	31	–	23
Social interests						
	0.028***	0.028***	0.042***	0.041***	0.042***	0.052***
	(0.002)	(0.003)	(0.003)	(0.004)	(0.010)	(0.011)
p	–	79	–	58	–	25
Importance of comfort aspects						
	-0.007***	-0.009***	-0.010***	-0.012***	-0.026***	-0.036***
	(0.002)	(0.003)	(0.003)	(0.004)	(0.009)	(0.010)
p	–	34	–	23	–	14
N	7098		4452		1616	

The table depicts the results of the effect of other factors about nurse's wage on the decision to become one. Standard errors are depicted in parentheses. Significance of the coefficient at conventional significance levels $1\%, 5\%, 10\%$ are indicated by stars ***, **, * respectively. N indicates the number of observations and p the number of chosen controls.

I examine the impact of the importance of economic factors on young people's involvement in nursing or perhaps even their withdrawal from nursing. The results are depicted in the first panel of table 5. Independent of the composition of the comparison group, I cannot conclude that the importance of economic factors plays a role in the decision to become a nurse. This result replicates findings of recent research: Nurses do not give much weight to economic factors. However, it is noticeable that future nurses do not weight economic factors lower than other individuals.

The next panel presents the results on the role of social interests in the decision to become a nurse. As expected, the results suggest that social interests play an important role in the decision to become a nurse. This holds true when the comparison group only consists of those who chose a social field. Compared to the effect of wage expectations, the effect of social interests is considerably larger (more than twice as large). The finding perfectly fits into both the nursing and the economic literature. It is often shown that preferences matter the most in the choice of training (e.g. Arcidiacono 2004, Wiswall & Zafar 2015). Therefore, the result provides some additional reassurance and further supports the results in table 3.

Nursing is generally known for its rather exhausting tasks and inflexible working hours. To investigate the effect of this reputation, I analyze the role of the importance comfort aspects on the probability of becoming a nurse. The results are presented in the last panel of table 5. They suggest that the larger the importance of comfort aspects, the lower the likelihood of becoming a nurse. Interestingly, compared to the coefficients in an unconditional model, the absolute size of the coefficients is larger in conditional models. That is, individuals that are more prone to become a nurse, put less emphasis on comfort aspects in their occupation. In summary, I find that the size of the effect of the expected wage is smaller than the role of individual interest, and that the effect of the expected wage has about the same size as the importance of comfort aspects.

4.4 Assessing wage information

As argued in section 4.1, the applied measures capture the information on wages rather than on expected wages (including own perceived ability). There may be three reasons for finding a positive effect of the expected wage. Results can be driven by future nurses who expect potentially overestimated wages or by non-future nurses that expect potentially underestimated wages. Moreover, both can occur simultaneously. To this end, I estimate further PDS models. Instead of the expected wage, I use measures that capture information relative to actual wages. As discussed above, changing the rank measure to information measures defined in equations (14)-(17) is straightforward. The results of the analysis are given in table 6.

In the first panel, I consider a measure that captures the general level of information defined in equation (14). The larger the measure, the higher the deviations from the actual relative wage and consequently the lower the level of information. The coefficient is standardized in order to assess its size. The result shows that an increase in the absolute cumulative deviation by one standard deviation decreases the probability to become a nurse by 0.6 percentage points. The effect becomes even larger when nurses are compared to those who chose a more similar occupation. This means that those who become a nurse can rank surveyed wages more precisely than those who do not become a nurse. Hence, I conclude that future nurses are well informed about relative wages.

In the remaining panels the effect of overestimation in equation (15), correct estimation in

Table 6 – Information about nurse's wages

	nurse vs. all		nurse vs. vocational training		nurse vs. social field	
	Single OLS	Post-Lasso	Single OLS	Post-Lasso	Single OLS	Post-Lasso
Cumulative absolute deviation to true ranks						
	-0.004**	-0.006***	-0.012***	-0.008**	-0.017**	-0.023**
	(0.002)	(0.002)	(0.003)	(0.003)	(0.008)	(0.009)
p	-	45	-	35	-	17
Nurse's wage rank overestimated						
	0.042***	0.033***	0.041***	0.038***	0.144***	0.111***
	(0.006)	(0.006)	(0.006)	(0.009)	(0.024)	(0.025)
p	-	40	-	17	-	18
Nurse's wage rank correctly estimated						
	0.007	0.002	0.010	0.002	0.008	0.008
	(0.005)	(0.005)	(0.008)	(0.008)	(0.020)	(0.019)
p	-	15	-	10	-	9
Nurse's wage rank underestimated						
	-0.026***	-0.018***	-0.032***	-0.024***	-0.082***	-0.061***
	(0.004)	(0.004)	(0.008)	(0.007)	(0.018)	(0.017)
p	-	27	-	13	-	9
N	7098		4452		1616	

The table depicts the results of the effect of information about nurse's wage on the decision to become one. The measures are defined in equations (14)-(17). The true ranking is: (1) barber, (2) motor vehicle mechanic, (3) nurse, (4) bank clerk, (5) teacher and (6) physician. The cumulative absolute deviation to true ranks is standardized. Standard errors are reported in parentheses. Significance of the coefficient at conventional significance levels $1\%, 5\%, 10\%$ are indicated by stars ***, **, * respectively. N indicates the number of observations and p the number of chosen controls.

equation (16) and underestimation in equation (17) on the probability to become a nurse is analyzed. The results show that overestimation of the nurse's rank increases and underestimation of the nurse's rank decreases the probability to become a nurse. Correct estimation does not affect the probability of becoming a nurse. These results remain statistically significant after controlling for an extensive set of confounders chosen by double-lasso-selection (e.g. general interests).

On the one hand, the results indicate that future nurses more often overestimate and less often underestimate the expected wage of a nurse. On the other hand, future nurses rank wages more in accordance with the true wages. Descriptive results in table 1 give further insights that are crucial for the interpretation of these seemingly contradicting results. The share of individuals who expect that nursing has the lowest wage among all six surveyed wages is much higher among non-future nurses than among future nurses (15% vs. 6%). In general, the share of individuals that underestimate the wage of a nurse is large (63%). Even a significant share of future nurses underestimate the wage (45%). In contrast, the share of those who overestimate a nurse's wage rank is low (13%). Furthermore, the comparison of mean expected wage ranks between future nurses and others in table 2 shows that there is only a significant difference in the expected rank of the three occupations with the lowest wages, i.e hairdresser, mechanic and nurse. There are no significant differences concerning occupations with higher wages, i.e. the expected wage rank of a bank clerk, teacher and physician. Therefore, I conclude that future nurses do not have exceptionally high wage expectations, but individuals who do not become a nurse have expectations that are too low. Even future nurses often expect relative wages that are too low. In summary, the analysis suggests that the perception of a low wage in nursing among young people may be an obstacle to attract more individuals to nursing.

5 Conclusion

This paper investigates the policy relevant question of whether and how wage information influences individual career choices to become a nurse. To this end, I used state-of-the-art methods for causal machine learning (post-double-selection, Belloni et al. 2014*a*) and sensitivity analysis (Cinelli & Hazlett 2020). My analysis does not use retrospective information that is potentially plagued by reverse causation, but longitudinal data following 9-th graders up to their decision whether or not to enter nursing training.

I report two sets of substantive findings. First, contrary to common perceptions, individuals' expectations about the wages in nursing do influence the probability of taking up nursing. The size of the effect is smaller than the effect of individual preferences but similar to other factors such as comfort aspects. Second, I show that understating the true rank of wages in nursing decreases the likelihood of starting a nursing career. My sensitivity analysis shows that potential unobserved confounders would have to be very strong to overrule these conclusions. The empirical results lead to two important policy implications. First, boosting wages in nursing may help to overcome the shortage observed in many countries. Second, providing more accurate information about actual (relative) wages in nursing would also help to attract more individuals into this profession.

References

Altonji, J. G., Arcidiacono, P. & Maurel, A. (2016), The analysis of field choice in college and graduate school: Determinants and wage effects, Vol. 5 of *Handbook of the Economics of Education*, Elsevier, pp. 305–396.

Andreassen, L., Di Tommaso, M. L. & Strøm, S. (2017), 'Nurses and physicians: a longitudinal analysis of mobility between jobs and labor supply', *Empirical Economics* **52**(4), 1235–1269.

Arcidiacono, P. (2004), 'Ability sorting and the returns to college major', *Journal of Econometrics* **121**(1-2), 343–375.

Beffy, M., Fougère, D. & Maurel, A. (2012), 'Choosing the field of study in postsecondary education: Do expected earnings matter?', *Review of Economics and Statistics* **94**(1), 334–347.

Belloni, A., Chen, D., Chernozhukov, V. & Hansen, C. (2012), 'Sparse models and methods for optimal instruments with an application to eminent domain', *Econometrica* **80**(6), 2369–2429.

Belloni, A., Chernozhukov, V. & Hansen, C. (2014*a*), 'High-dimensional methods and inference on structural and treatment effects', *Journal of Economic Perspectives* **28**(2), 29–50.

Belloni, A., Chernozhukov, V. & Hansen, C. (2014*b*), 'Inference on treatment effects after selection among high-dimensional controls', *Review of Economics and Statistics* **81**(2), 608–650.

Blossfeld, H.-P. & von Maurice, J. (2011), 'Education as a lifelong process', *Zeitschrift für Erziehungswissenschaft* **14**(S2), 19–34.

Bomball, J., Schwanke, A., Stöver, M., Schmitt, S. & Görres, S. (2010), 'Imagekampagne für Pflegeberufe auf der Grundlage empirisch gesicherter Daten: Einstellungen von Schüler/innen zur möglichen Ergreifung eines Pflegeberufes'.
URL: *http://www.ipp.uni-bremen.de/uploads/IPPSchriften/*

Boudarbat, B. (2008), 'Field of study choice by community college students in Canada', *Economics of Education Review* **27**(1), 79–93.

Brunello, G., Lucifora, C. & Winter-Ebmer, R. (2004), 'The wage expectations of European business and economics students', *Journal of Human Resources* **XXXIX**(4), 1116–1142.

Carneiro, P., Heckman, J. J. & Vytlacil, E. (2011), 'Estimating marginal returns to education', *The American economic review* **101**(6), 2754–2781.

Cho, S.-H., Jung, S. Y. & Jang, S. (2010), 'Who enters nursing schools and why do they choose nursing? A comparison with female non-nursing students using longitudinal data', *Nurse education today* **30**(2), 180–186.

Cinelli, C. & Hazlett, C. (2020), 'Making sense of sensitivity: Extending omitted variable bias', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **82**(1), 39–67.

Doiron, D., Hall, J., Kenny, P. & Street, D. J. (2014), 'Job preferences of students and new graduates in nursing', *Applied Economics* **46**(9), 924–939.

Fernandes, A., Huber, M. & Vaccaro, G. (2020), 'Gender differences in wage expectations'.
URL: *http://arxiv.org/pdf/2003.11496v1*

German Employment Agency (2018), 'Entgeltatlas'.
URL: *https://entgeltatlas.arbeitsagentur.de/entgeltatlas*

German Employment Agency (2020), Arbeitsmarkt im Pflegebereich, Berichte: Blickpunkt Arbeitsmarkt 05/2020.

Hanel, B., Kalb, G. & Scott, A. (2014), 'Nurses' labour supply elasticities: the importance of accounting for extensive margins', *Journal of health economics* **33**, 94–112.

Hastie, T., Friedman, J. & Tibshirani, R. (2009), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer Series in Statistics, second edn, Springer-Verlag New York, New York, NY.

Knoll, B., Riedel, N. & Schlenker, E. (2017), 'He's a chip off the old block - the persistence of occupational choices across generations', *Labour* **31**(2), 174–203.

Matthes, S. (2019), *Warum werden Berufe nicht gewählt? Die Relevanz von Attraktions- und Aversionsfaktoren in der Berufsfindung*, Berichte zur beruflichen Bildung, 1. auflage edn, Bundesinstitut für Berufsbildung, Bonn.

McCabe, R., Nowak, M. & Mullen, S. (2005), 'Nursing careers: What motivated nurses to choose their profession?', *Australian Bulletin of Labour* **31**(4), 380–402.

Montmarquette, C., Cannings, K. & Mahseredjian, S. (2002), 'How do young people choose college majors?', *Economics of Education Review* **21**(6), 543–556.

Mullainathan, S. & Spiess, J. (2017), 'Machine learning: An applied econometric approach', *Journal of Economic Perspectives* **31**(2), 87–106.

Oster, E. (2019), 'Unobservable selection and coefficient stability: Theory and evidence', *Journal of Business & Economic Statistics* **37**(2), 187–204.

Petrucci, C., La Cerra, C., Aloisio, F., Montanari, P. & Lancia, L. (2016), 'Empathy in health professional students: A comparative cross-sectional study', *Nurse education today* **41**, 1–5.

Schweri, J. & Hartog, J. (2017), 'Do wage expectations predict college enrollment? Evidence from healthcare', *Journal of Economic Behavior & Organization* **141**, 135–150.

Shields, M. A. (2004), 'Addressing nurse shortages: What can policy makers learn from the econometric evidence on nurse labour supply?', *The Economic Journal* **114**(499), F464–F498.

Speer, J. D. (2019), 'Where the girls are: Examining and explaining the gender gap in the nursing major', *Scottish Journal of Political Economy* **121**(4), 1.

Tibshirani, R. (1996), 'Regression shrinkage and selection via the lasso', *Journal of the Royal Statistical Society: Series B (Methodological)* **58**(1), 267–288.

Urminsky, O., Hansen, C. & Chernozhukov, V. (2016), 'Using double-lasso regression for principled variable selection', *SSRN Electronic Journal* .

van Buuren, S. & Groothuis-Oudshoorn, K. (2011), 'mice: Multivariate imputation by chained equations in R', *Journal of Statistical Software* **45**(3).

Wilkes, L., Cowin, L. & Johnson, M. (2015), 'The reasons students choose to undertake a nursing degree', *Collegian (Royal College of Nursing, Australia)* **22**(3), 259–265.

Willis, R. J. & Rosen, S. (1979), 'Education and self-selection', *Journal of Political Economy* **87**(5, Part 2), S7–S36.

Wiswall, M. & Zafar, B. (2015), 'Determinants of college major choice: Identification using an information experiment', *Economics of Education Review* **82**(2), 791–824.

Wohlkinger, F., Ditton, H., von Maurice, J., Haugwitz, M. & Blossfeld, H.-P. (2011), 'Motivational concepts and personality aspects across the life course', *Zeitschrift für Erziehungswissenschaft* **14**(2), 155.

A Tables

Table A1 – Summary statistics of potential controls

Variable	Mean	Standard Deviation	Min	Max	Number of Non-Missing Values
Demographic					
Gender: female	0.4978	0.5000	0.0	1.0	7089
Migration background	0.1035	0.3047	0.0	1.0	7089
Opportunities					
Higher secondary track in 9th grade	0.3957	0.4890	0.0	1.0	6907
Competencies					
Science	0.1561	0.9793	-2.6	5.3	6985
Mathematics	0.2077	1.2220	-4.4	4.6	7002
Information and communication technology	0.1461	0.9002	-3.3	4.1	6989
Reading	34.6639	8.3705	0.0	51.0	7003
Reading speed	0.1577	1.2019	-4.0	3.3	6950
Metacognition	0.8185	0.1175	0.0	1.0	6965
Attitude to school & school performance					
Math grade: over average (class)	0.5445	0.4981	0.0	1.0	6981
German grade: over average (class)	0.5083	0.5000	0.0	1.0	7016
Grade: math	2.8770	0.9989	1.0	6.0	6981
Grade: german	2.8077	0.7970	1.0	6.0	7016
Ever retent a grade	0.1534	0.3604	0.0	1.0	6984
School concept: german	2.9434	0.6233	1.0	4.0	7022
School concept: math	2.5750	0.9186	1.0	4.0	6993
School concept: general	2.9264	0.5630	1.0	4.0	7019
Interests in math	2.2180	0.7900	1.0	4.0	6862
Interests in german	2.1874	0.8047	1.0	4.0	6865
Personality & behavior					
Big Five: artistic	2.7984	1.3196	1.0	5.0	7070
Big Five: crticize	2.8645	1.0243	1.0	5.0	7068
Big Five: easy-going/lazy	3.2171	1.1723	1.0	5.0	7074
Big Five: nervous	2.8271	1.0795	1.0	5.0	7066
Big Five: imaginative	3.7465	1.0257	1.0	5.0	7062
Big Five: relaxed	3.2995	1.0612	1.0	5.0	7072

Table A1 continued from previous page

Big Five: cautious/relaxed	2.6416	1.1134	1.0	5.0	7071
Big Five: sensitive	3.8503	0.9112	1.0	5.0	7068
Big Five: sociable	3.5319	0.9446	1.0	5.0	7061
Big Five: thorough	3.5920	0.9284	1.0	5.0	7071
Big Five: trusting	3.3983	1.0098	1.0	5.0	7068
Considerate	2.6108	0.5218	1.0	3.0	7011
Gets mobbed	1.1665	0.4243	1.0	3.0	6980
Has friends	2.8863	0.3434	1.0	3.0	7000
Helpful	2.6887	0.4996	1.0	3.0	6997
Kind to younger	2.4744	0.5924	1.0	3.0	6992
Likes to help	2.2183	0.5872	1.0	3.0	6994
Loner	1.4503	0.6032	1.0	3.0	6986
Popular	2.3546	0.5770	1.0	3.0	6940
Global self-esteem	2.4496	1.1497	1.0	5.0	7004
Likes to share	2.5696	0.5503	1.0	3.0	7005
Gets along with adults	1.6721	0.6596	1.0	3.0	6990
Good as others	3.9471	0.7928	1.0	5.0	7068
Be a failure	1.6985	0.9576	1.0	5.0	7045
Good qualities	3.9625	0.7622	1.0	5.0	7060
No pride	2.0150	0.9865	1.0	5.0	7059
Positive attitude towards myself	3.9402	0.9088	1.0	5.0	7056
Satisfied with myself	3.9442	0.8371	1.0	5.0	7078
No good	2.3150	1.0705	1.0	5.0	7053
Feel useless	1.8723	1.0089	1.0	5.0	7057
Be at least as valuable as others	4.0033	0.9849	1.0	5.0	7054
TenFlex: flexible	16.0394	3.2374	5.0	25.0	7007
TenFlex: persistent	18.4130	2.9034	5.0	25.0	7063
Religious	2.2474	0.8976	1.0	4.0	6810
Disadvantage: gender	0.0740	0.2617	0.0	1.0	6341
Disadvantage: foreign name	0.3457	0.4756	0.0	1.0	6411
Disadvantage: foreign looks	0.3505	0.4772	0.0	1.0	6400
Disadvantage: lower secondary	0.7906	0.4069	0.0	1.0	6620
Disadvantage: head scarf	0.5563	0.4969	0.0	1.0	6052
Disadvantage: overweight	0.2038	0.4028	0.0	1.0	6370

Table A1 continued from previous page

Disadvantage: bad german	0.8574	0.3497	0.0	1.0	6583
Family & career planning					
Important to form family	0.6812	0.4661	0.0	1.0	7085
Child before age 25	0.2307	0.4213	0.0	1.0	7079
Moving away for training	0.4160	0.4929	0.0	1.0	6019
Satisfaction					
Satisfaction with life	7.5462	1.9546	0.0	10.0	7089
Satisfaction with living conditions	8.0968	1.8901	0.0	10.0	7089
Satisfaction with family	8.3861	2.1703	0.0	10.0	7089
Satisfaction with friends	8.6148	1.8390	0.0	10.0	7089
Satisfaction with school	6.8558	2.2203	0.0	10.0	7089
Satisfaction with health	8.3168	2.0770	0.0	10.0	7089
Leisure					
Time gaming	3.0198	1.5115	1.0	6.0	6910
Time reading	3.1293	1.4694	1.0	5.0	6927
Visiting museum	2.2284	1.0784	1.0	5.0	7057
TV-shows: science	1.9890	0.7473	1.0	4.0	7024
Books: science	1.4038	0.6478	1.0	4.0	7024
Web: science	1.7473	0.7735	1.0	4.0	7012
Magazines: science	1.7091	0.7909	1.0	4.0	7013
Science club	1.1463	0.4802	1.0	4.0	7020
Course: music	1.7950	0.4037	1.0	2.0	7089
Number of books	3.9537	1.4359	1.0	6.0	7064
Meaning of work and interests					
Importance of comfort aspects	4.6524	0.9508	1.0	6.0	7089
Importance of economic aspects	5.1635	0.7465	1.0	6.0	7089
Importance of expressive aspects	4.9322	0.6508	1.0	6.0	7052
IILS-Interests: social	3.0449	0.9829	1.0	5.0	7089
IILS-Interests: conventional	2.5018	0.8550	1.0	5.0	7057
IILS-Interests: art	2.5329	1.0149	1.0	5.0	7065
IILS-Interests: analytical	2.6614	0.9723	1.0	5.0	7076
IILS-Interersts: practical	2.8324	1.0586	1.0	5.0	7066
IILS-Interests: business	3.0338	0.8357	1.0	5.0	7060
Parental background					
Parental education (highest): studied	0.2918	0.4546	0.0	1.0	5452

Table A1 continued from previous page

Parental education (highest): university entrance quali.	0.1970	0.3978	0.0	1.0	5452
Household income per capita	859.8624	392.4184	200.0	2666.7	4198
Parental occupation (at least one parent): MINT	0.5144	0.4998	0.0	1.0	5476
Parental occupation (at least one parent): business	0.5338	0.4989	0.0	1.0	5476
Parental occupation (at least one parent): care	0.0942	0.2922	0.0	1.0	5476
Parental occupation (at least one parent): health	0.1348	0.3415	0.0	1.0	5476
Parental occupation (at least one parent): education	0.1715	0.3770	0.0	1.0	5476
Parental occupation (at least one parent): hairdresser	0.0197	0.1391	0.0	1.0	5476
Parental occupation (at least one parent): banking	0.0499	0.2177	0.0	1.0	5476
Parental occupation (at least one parent): automotive mechanic	0.0268	0.1616	0.0	1.0	5476
Parental occupation (at least one parent): teacher	0.0575	0.2329	0.0	1.0	5476
Parental occupation (at least one parent): physician	0.0247	0.1551	0.0	1.0	5476
Broken home	0.0900	0.2862	0.0	1.0	6832
Behavior and values of parents					
Discuss books	1.8053	1.0344	1.0	5.0	6930
Discuss movies	3.2535	1.1332	1.0	5.0	6928
Discuss politcs	2.5871	1.2831	1.0	5.0	6943
Discuss arts	1.5531	0.9378	1.0	5.0	6950
Importance to maintain mother's status (education)	3.6373	1.3416	1.0	5.0	6515
Importance to maintain father's status (education)	3.6483	1.3606	1.0	5.0	6381
Importance of grads	4.3354	0.8863	1.0	6.0	7015
Importance of parent's opinion	3.9303	0.9589	1.0	5.0	7013
Gender role: duties in household	3.2981	0.8462	1.0	4.0	6032
Gender role: technology	2.7174	0.8996	1.0	4.0	5937

Table A1 continued from previous page

Gender role: politics	3.2233	0.8659	1.0	4.0	6883
Gender role: earning money	1.8850	0.9479	1.0	4.0	6992
Gender role: occupations	3.0044	0.9044	1.0	4.0	6991
Importance career	4.0437	1.1019	0.0	5.0	7004
Importance to maintain mothers status (occupation)	3.7709	1.2631	1.0	5.0	6966
Importance to maintain fathers status (occupation)	3.7208	1.2600	1.0	5.0	6923
Expectations of son: living close	2.0428	0.7851	1.0	4.0	6334
Expectations of son: housekeeping	2.5314	0.9255	1.0	4.0	6536
Expectations of son: financially support younger siblings	1.8419	0.8128	1.0	4.0	6235
Expectations of daughter: living close	2.2826	0.8827	1.0	4.0	6295
Expectations of daughter: housekeeping	2.8661	0.8996	1.0	4.0	6475
Expectations of daughter: financially support younger siblings	1.8036	0.7886	1.0	4.0	6104
Expectations to study	0.4272	0.4947	0.0	1.0	6645
Costs of lower secondary degree	3.3527	1.1535	1.0	5.0	6941
Costs of middle secondary degree	3.7966	0.8572	1.0	5.0	6914
Costs of higher secondary degree	3.9547	1.0164	1.0	5.0	6908
Social environment					
Organization: sports	0.6558	0.4751	0.0	1.0	6990
Organization: religion	0.2136	0.4099	0.0	1.0	6946
Organization: culture	0.1434	0.3505	0.0	1.0	6932
Friends: share migration background	2.6436	1.3255	1.0	7.0	7085
Friends: share ambitious	3.1854	0.7628	1.0	5.0	7012
Friends: share try	2.7581	1.0162	1.0	5.0	7068
Friends: share don't care	2.5054	0.9814	1.0	5.0	7012
Friends: important to have a career	3.6128	0.8680	1.0	5.0	7066
Class: share migration background	2.6724	1.1206	1.1206	7.01	6979
Class: share ambitious	3.0810	0.7637	1.0	5.0	6999
Class: share try	2.4859	0.9372	1.0	5.0	7065

Table A1 continued from previous page

Class: share don't care	2.7202	0.9269	1.0	5.0	6987
School					
Teacher: further education about voc. orientation	3.1783	0.9487	1.0	5.0	5681
Contact: organization	3.6825	0.9318	1.0	5.0	5603
Contact: firms	3.9968	0.9016	1.0	5.0	5642
Programs for voc. orientation	4.1166	0.9513	1.0	5.0	5704
Contact: counseling	3.8845	0.9966	1.0	5.0	5669
Contact: local network	3.6833	1.1447	1.0	5.0	5671
Parental support in voc. orientation	3.6765	0.9647	1.0	5.0	5685
Testing of interests	4.2433	1.7749	1.0	6.0	5474
Individual support plans	2.5473	1.5010	1.0	6.0	5375
Voc. orientation by teachers	4.9588	1.6463	1.0	6.0	5458
Practice: writing applications	5.7043	0.9610	1.0	6.0	5523
Practice: job interview	5.1684	1.4346	1.0	6.0	5468
Train social competencies	4.3466	1.8184	1.0	6.0	5366
Assisted internship	5.1491	1.6617	1.0	6.0	5487
External counseling	4.6583	1.7494	1.0	6.0	5502
Voc. orientation in institutions	3.4604	1.9691	1.0	6.0	5426
Individual counseling	3.1498	1.6759	1.0	6.0	5499
Individual support by career choice assistance	2.2650	1.5128	1.0	6.0	5399
Support by educators	2.0501	1.1373	1.0	6.0	5472
Regional and labor market characteristics					
Share age 15 to 25	11.7035	0.7721	9.9	14.4	7072
Firm density	42.0207	39.1636	2.4	186.7	7072
Regional unemployment rate	2.4753	1.0799	1.0	4.0	7072
Residence in east-germany	0.1251	0.3308	0.0	1.0	6835

IAW-Diskussionspapiere

Die IAW-Diskussionspapiere erscheinen seit September 2001. Die vollständige Liste der IAW-Diskussionspapiere von 2001 bis 2013 (Nr. 1-110) finden Sie auf der IAW-Internetseite http://www.iaw.edu/index.php/IAW-Diskussions-papiere.

IAW-Diskussionspapiere ab 2013:

Nr. 111 (Oktober 2014)
Offshoring Potential and Employment Dynamics
Bernhard Boockmann

Nr. 112 (Oktober 2014)
Is Offshoring Linked to Offshoring Potentials? Evidence from German Linked-Employer-Employee Data
Tobias Brändle

Nr. 113 (November 2014)
University Knowledge and Firm Innovation – Evidence from European Countries
Andrea Bellucci / Luca Pennacchio

Nr. 114 (Januar 2015)
We Want them all Covered! Collective Bargaining and Firm Heterogeneity. Theory and Evidence from Germany
Florian Baumann / Tobias Brändle

Nr. 115 (Januar 2015)
Coaching, Counseling, Case-Working: Do They Help the Older Unemployed out of Benefit Receipt and back into the Labor Market?
Bernhard Boockmann / Tobias Brändle

Nr. 116 (Januar 2015)
The One Constant: A Causal Effect of Collective Bargaining on Employment Growth? Evidence from German Linked-Employer-Employee Data
Tobias Brändle / Laszlo Goerke

Nr. 117 (Februar 2015)
Activation as a First Step: Estimating the Causal Effects of a Job Search Assistance Programme
Tobias Brändle / Lukas Fervers / Carina Webel

Nr. 118 (März 2015)
Mental Representation of Sharing Experiments: Analyzing Choice and Belief Data
Werner Güth /Charlotte Klempt / Kerstin Pull

Nr. 119 (April 2015)
Collateral and Local Lending: Testing the Lender-Based Theory
Andrea Bellucci / Alexander Borisov / Germana Giombini / Alberto Zazzaro

Nr. 120 (April 2015)
The Great Trade Collapse and the Spanish Export Miracle: Firm-level Evidence from the Crisis
Peter S. Eppinger / Nicole Meythaler / Marc-Manuel Sindlinger / Marcel Smolka

Nr. 121 (Juni 2015)
Do exporting firms benefit from retail internationalization?
Angela Cheptea / Charlotte Emlinger / Karine Latouche

Nr. 122 (Januar 2016)
Bank Organization and Loan Contracting in Small Business Financing
Andrea Bellucci / Alexander Borisov/ Alberto Zazzaro

Nr. 123 (März 2016)
Mentoring Disadvantaged Youths during School-to-Work Transition: Evidence from Germany
Bernhard Boockmann / Sebastian Nielen

IAW-Diskussionspapiere

Nr. 124 (Mai 2016)
Ökonomische Wirkungen der Handwerksnovelle 2004: Ergebnisse aus einem quasinatürlichen Experiment
Andreas Koch / Sebastian Nielen

Nr. 125 (Juni 2016)
Fast track to the labour market or highway to hell? The effect of activation policies on quantity and quality of labour market integration
Lukas Fervers

Nr. 126 (Oktober 2016)
Handelsliberalisierung: Unilateral, regional, multilateral
Wilhelm Kohler

Nr. 127 (Oktober 2016)
Trade creating oder Trade diverting – Ökonomische Perspektiven auf den Abbau technischer Handelshemmnisse in multilateralem oder regionalem Rahmen
Benjamin Jung

Nr. 128 (November 2016)
The Distribution of Household Savings in Germany
Jochen Späth / Kai Daniel Schmid

Nr. 129 (April 2017)
Give it Another Try: What are the Effects of a Public Employment Scheme Especially Designed for Hard-to-Place Workers?
Tobias Brändle / Lukas Fervers

Nr. 130 (Januar 2018)
Integration der Geflüchteten in den Arbeitsmarkt: Ziele, Strategien und Hemmnisse aus Sicht der Jobcenter
Bernhard Boockmann / Tobias Scheu

Nr. 131
Operationalizing Seasonal Work in Germany (September 2018)
Jochen Späth / Tobias Brändle / Stefan Preuss / Marcel Reiner

Nr. 132
Rather a trigger than a cause of change. Responses of firms and workers to the statutory minimum wage in Germany (Oktober 2018)
Andreas Koch / Andrea Kirchmann / Marcel Reiner / Tobias Scheu / Holger Bonin

Nr. 133
Accounting for Investment Risk in Educational Decisions: New Evidence for Lifetime Returns in Germany (April 2020)
Anne Zühlke / Philipp Kugler / Armin Hackenberger / Tobias Brändle

Nr. 134 (Dezember 2020)
The Attractiveness of Jobs in the German Care Sector – Results of a Factorial Survey
Martin Kroczek / Jochen Späth

Nr. 135 (Januar 2021)
Wage expectation, information and the decision to become a nurse
Philipp Kugler