Variable Selection with Group Structure

Exiting Employment at Retirement Age - A Competing Risks Quantile Regression Analysis

Shi, Shuolin ; Wilke, Ralf A.

Document Version
Accepted author manuscript

Published in:
Empirical Economics

DOI:
10.1007/s00181-020-01918-z

Publication date:
2022

License
Unspecified

Citation for published version (APA):
Shi, S., & Wilke, R. A. (2022). Variable Selection with Group Structure: Exiting Employment at Retirement Age - A Competing Risks Quantile Regression Analysis. *Empirical Economics*, *62*(1), 119-155. https://doi.org/10.1007/s00181-020-01918-z

Variable Selection with Group Structure: Exiting employment at retirement age - A Competing Risks Quantile Regression Analysis[1]

Shuolin Shi, Copenhagen Business School, Department of Economics, ssh.eco@cbs.dk
Ralf A. Wilke, Copenhagen Business School, Department of Economics, rw.eco@cbs.dk

Abstract We consider the exit routes of older employees out of employment around retirement age. Our administrative data cover weekly information about the Danish population from 2004 to 2016 and 397 variables from 16 linked administrative registers. We use a flexible dependent competing risks quantile regression model to identify how early and late retirement transitions are related to the information in the various registers. Our model selection is guided by machine learning methods, in particular statistical regularization. We use the (adaptive) group bridge to identify the relevant administrative registers and variables in heterogeneous and high dimensional data, while maintaining the oracle property. By applying state-of-the-art statistical methods, we obtain detailed insights into conditional distributions of transition times into the main pension programs in Denmark.

Keywords Adaptive group bridge, Competing risks, Quantile regression, Statistical learning

JEL Classification C55, J26

1 Introduction

In an attempt to make the pension system fit for the future, politics in Denmark introduced more flexibility on the timing of retirement during the 2000s. This resulted for the employed in the possibility to decide on the retirement point once a certain age threshold has been passed. The point of retirement is then no longer deterministic, but varies across individuals and is likely dependent on a wealth of individual, economic and institutional factors. Previous analysis has mainly studied early retirement patterns. In this paper, we consider both early and late retirement. The event to withdraw from employment may be in the discretion of the employed but can be also due to factors out of her control such as invalidity or dismissal. To account for this complexity we choose a competing risks duration model, where each risk corresponds to a different exit route into retirement. Because we expect early and late retirement pattern to differ, we require a flexible model for our analysis that allows the determinants of exiting employment to affect the conditional distributions differently for long and short durations. It is well known that quantile regression imposes milder restrictions on the covariate effects than conventional models such as proportional hazards (Koenker and Geling, 2001). While still relatively rarely used in economic application with survival data (e.g. Koenker and Bilias, 2001, Fitzenberger and Wilke, 2010a), the method has enjoyed increasing popularity in biostatistics and a number of practical model extensions have been developed which are summarized in Koenker et al. (2017). Competing risks duration models

[1] We thank an associate editor and a reviewer for helpful comments, the Pension Research Centre (Percent) for financial support and Kwang Woo Ahn for making sample code for the (adaptive) group bridge available.

are not identifiable without additional restrictions (Cox, 1962) and resulting identification bounds are typically wide and uninformative in applications (Peterson, 1976). The model becomes identifiable when imposing restrictions on the covariate structure and the role of the covariates (Heckman and Honoré, 1989), although these restrictions can be hardly verified in practice (Femanian, 2003). What can be identified and estimated without strong restrictions are cumulative incidences and subdistribution functions as these relate to the distribution of observed transitions. For the link between the models for the different hazards and distributions, compare Emura et. al. (2019). In our analysis, we adopt the competing risks quantile regression model for the cumulative incidence (Peng and Fine, 2009). Given our data structure with several hundred covariates, we combine the quantile regression approach with statistical regularization techniques to obtain a statistical learning based selection of the factors that make people leave their job earlier or later. These techniques possess the oracle property and therefore have desirable statistical properties.

There is an extensive literature that considers transitions out of employment into (early) retirement (e.g. Lindeboom, 1998, Duval, 2003). Motivated by the ageing of the societies and subsequent restraints for the financial situation of the pension funds (Gruber and Wise, 1998), the question is analyzed how the institutional system can be shaped in order to avoid incentives to retire early. Beside direct (early) retirement, people may choose other forms of exit routes, such as exits through a bridging period in unemployment or disability (see e.g. Miniaci and Stancanelli, 1998, Kyyrä and Wilke, 2007, Fitzenberger and Wilke, 2010b, Bingley et al., 2012). Relevant literature for Denmark has considered general determinants of retirement (e.g. Filges et al., 2012, Larsen and Pedersen, 2013, Kallestrup-Lamb et al., 2016). The use of duration models is limited to a couple of studies (An et al., 2004, Christensen and Kallestrup-Lamb, 2012, Gørtz, 2012) and based on annual data for years before 2008. Therefore, these analyses are about periods when a less flexible system was in place. Due to the low frequency of the annual data only discrete time or discrete choice models have been applied.

Our analysis is based on weekly, monthly and annual observations of employees in Denmark for the period 2004-2016. It uses administrative data provided by Statistics Denmark that links 16 different registers. We therefore face a complex data set with numerous partly highly collinear variables and group structure. In the context of our application, we define an administrative register as a group of variables as each register contains a set of variables on related statistics such as health, crime or labor market, and one variable can enter more than one groups. We use variants of the Lasso (Tibshirani, 1996) that accommodate these features of the data structure. In particular, we use group level and bi-level variable selection methods for competing risks quantile regression (Ahn and Kim, 2018). Developed for problems in medical sciences and tested with data on gene selection, these methods identify blocks of genes and individual genes that are in relation with diseases. We explore how these methods perform with our more heterogeneous linked administrative social sciences data. The use of the Lasso in economic problems is still not widespread but increasing, though most applications are for commonly used mean regression or discrete choice models and do not consider group or bi-level variable selection. Our quantile regression model is more complex as it is estimated separately for different quantiles. Therefore, the resulting set of variables changes by quantile. As the adaptive group bridge permits consistent identification of non-zero groups and within-group variables, we explore how it is helpful in identifying relevant registers and within-register variables for the economic problem at hand. Our approach is therefore not only of interest to empirical researchers how to include the relevant variables but also a relevant tool for data providers, as researchers should be only given access to the minimal set of relevant administrative registers due to data protection regulations.

Our study contributes to the literature as follows: It is the first application of these methods to large-scale

administrative data. We are not aware that the (adaptive) group bridge has been applied in social science literature, economics in particular, or in combination with competing risks quantile regression for survival analysis. We therefore explore the practical properties of combining a flexible and complex distributional model with state of the art statistical regularization methods. For the analysis of Denmark, we contribute by using weekly data until the year 2016 to conduct a dependent competing risks duration analysis. Our analysis period is characterized by a more flexible retirement system, which permits us to study early and late retirement in one model.

The rest of the paper is organized as follows. In Section 2, we briefly review the existing literature about determinants of retirement in Denmark. Section 3 gives an overview of the main retirement programs in Denmark. In Section 4, we describe the dataset. Section 5 briefly reviews the research methods. In Section 6, we present and analyze the empirical results. In Section 7, we give conclusions and discussions.

2 Literature Review

We briefly review the literature on transitions to (early) retirement in Denmark. Various studies have considered general determinants of retirement (e.g. Filges et al., 2012, Larsen and Pedersen, 2013, Kallestrup-Lamb et al., 2016), while others focus on a specific determinant (e.g. Danø et al., 2005, Christensen and Kallestrup-Lamb, 2012). Most studies focus on individual retirement, while some focus on joint retirement of married couples (e.g. An et al., 2004, Bingley and Lanot, 2007). Some studies focus on pathways to (early) retirement (Larsen and Pedersen, 2005, Bingley et al., 2012), while others focus on late retirement (e.g. Amilon and Nielsen, 2010) and semi-retirement (e.g. Larsen and Pedersen, 2013). Most studies use discrete response models (e.g. Filges et al., 2012, Larsen and Pedersen, 2013, Bingley et al., 2016, Kallestrup-Lamb et al., 2016) or failure time models (e.g. An et al., 2004, Christensen and Kallestrup-Lamb, 2012), while few studies focus on aggregate statistics (e.g. Barslund, 2015, OECD, 2012a). Hardly any of the existing studies use data for years after 2008 and therefore they rely on annual data.

The uses of duration models include An et al. (2004) and Christensen and Kallestrup-Lamb (2012). An et al. (2004) study the joint retirement decisions of Danish married couples. Specifically, they examine whether the retirement timing of married couple is determined individually or jointly. Results show that financial and health variables play significant roles in explaining both individual retirement and joint retirement decisions; complementarities in leisure time explain joint early retirement decisions; correlation in unobserved heterogeneity, such as common tastes, plays a larger role than other observed heterogeneity in explaining joint late retirement decisions. Overall, retirement is a household decision. Christensen and Kallestrup-Lamb (2012) study the determinants of duration until retirement, in particular the impact of changes in health status on early retirement behavior. Their study is based on annual panel data for working people from 1985 to 2001. Results show that health and other regressors have different effects on disability retirement, early retirement, unemployment followed by early retirement and by other programs. Gørtz (2012) uses a discrete-time proportional hazard model to study the early retirement behavior of female teachers in the day-care sector.

Another interesting study by Kallestrup-Lamb et al (2016) focuses on the general determinants of retirement using the adaptive Lasso applied to logistic regression. The study is based on annual data for working people for the year 1980 and 1998 and is the first application of Lasso-type estimator to this type and scale of data. Their dataset includes 399 variables covering demographic, socioeconomic, financial, health, labor market status, lags of time-varying regressors, and characteristics of the spouse if the individual is married. All types

of retirement are pooled. The penalized logistic regression model uses both the Logit and the Lasso estimator as an initial estimator. Results show that the choice of initial estimator for adaptive Lasso matters in terms of the number of selected variables. Their analysis suggests that Lasso-type estimators produce reasonable results for the variable selection problem at hand.

3 Institutional Setup

Denmark as most European countries has introduced more flexibility on the timing of retirement during the 2000s in order to motivate working longer through, e.g. semi-retirement or late retirement. In this section, we give an overview of the main retirement routes through public pension or labor market pension systems, as well as their main changes in the 2000s related to extending working life.

Old age pension (OAP), or state pension, is a universal pension that applies to every Danish national who has lived in Denmark for at least three years between the age of 15 and 65, and aims to protect the elderly from poverty. Besides the pension itself, pensioners may be eligible for different benefits, such as housing benefit, heating benefit, health-related benefit, etc. The retirement age has gone through several changes. For people born before 1 July 1939, the retirement age is 67; for people born between 1 July 1939 and 1 January 1963, the retirement age is shown in Table 1; for people born after 1 January 1963, the retirement age is according to future life expectancy.

The old age pension consists of a basic amount and a pension supplement and is means-tested. Although, the test against income was reduced in the 2006 welfare reform, it is still unattractive to work and receive the state pension at the same time, as there is no or a reduced entitlement for higher income levels. From 1 July 2004, a pension deferral policy was introduced to motivate people to continue working after retirement age. People can postpone the state pension and get a higher payment afterwards if they work for at least 750 hours a year in the deferral period. Note that before 2011, the qualifying working hours were 1,500 in 2004 and 1,000 in 2008. The maximum deferral period is ten years and people can defer the pension for two times.

Table 1 Retirement Age of Old Age Pension and Early Retirement Pension

Date of Birth	Old Age Pension	Early Retirement Pension
1 Jul 1939 – 31 Dec 1953	65	60
1 Jan 1954 – 30 Jun 1954	65.5	60.5
1 Jul 1954 – 31 Dec 1954	66	61
1 Jan 1955 – 30 Jun 1955	66.5	61.5
1 Jul 1955 – 31 Dec 1955	67	62
1 Jan 1956 – 30 Jun 1956	67	62.5
1 Jul 1956 – 31 Dec 1958	67	63
1 Jan 1959 – 30 Jun 1959	67	63.5
1 Jul 1959 – 31 Dec 1962	67	64
1 Jan 1963 –	68	65

Source: Borger.dk (2019)

Disability pension, or early Danish pension, is another universal pension in Denmark. It applies to Danish nationals who are between the ages of 18 and 65, have lived in Denmark for at least three years between the age of 15 and 65, and meet reduced work capacity criteria. After a disability pension reform in June 2012, the

minimal age for disability pension was increased to 40 and people under 40 years old can only receive disability pension under special circumstances. Pensioners receive a certain amount depending on income level and receive some other housing and healthcare benefits.

Early retirement pension, or post-employment wage (PEW) program, is a voluntary labor market pension. People can choose to be fully insured or partially insured. The scheme was introduced in 1979 in order to balance the unemployment of young people and the employment of older people. Pensioners have the opportunity to retire before the state retirement age and maintain a decent income level. Eligibility requires membership of an unemployment insurance fund, continuous contributions for at least 30 years, employment higher than 1,924 working hours or income higher than 233,375 DKK within the last three years, and residence in Denmark. Similar to the state pension, the minimum retirement age for early retirement pension has changed several times as shown in Table 1. For people born before 1 January 1954, the early retirement age is 60, which means that the maximum duration of early retirement is 5 years; for people born later, the early retirement age gradually increases to 65 and the duration gradually reduces to 3 years.

The payment of early retirement pension differs according to previous income and insurance level, and reduces if the pensioner has income from labor market pension, individual pension, work, etc. The maximum payment is the minimum of 90% of previous income and 91% of the unemployment insurance benefit. For people born before July 1 1959, they can choose to postpone pension by working for at least 1,560 hours per year for fully insured and 1,248 hours for partially insured, and receiving early retirement pension no more than three years before the state retirement age. By postponing, they can get 100% of the unemployment insurance benefit and a tax-free premium for wages. Those born before 1 January 1956 can additionally earn a set-off amount for other pension income.

Other exit routes to early retirement include civil servants' pension, partial pension. among others. We only consider the above-mentioned three retirement programs, as they are the main exit routes to retirement and due to data availability.

4 Data

We use register data from Statistics Denmark (DST) and the DREAM database, which contain weekly, monthly and annual observations for the population in the period 2004-2016. We use 397 variables from 16 registers of DST's linked administrative data as regressors, and use DREAM to generate competing risks and durations. In our analysis, we restrict the sample to individuals born in 1949 who have stable employment history from 2004 to 2008, and follow their employment and retirement status from 2009 to 2016. This leaves us with 8,178 individuals who have not shown limited work ability or long-term illness before the age of 59. More details on the construction of our sample are given in Supplementary Material S.3. We consider two main exit routes from employment: 1. retirement (via disability pension, early retirement pension and old age pension); 2. other exits (via unemployment, illness, death, etc.). These routes define the two competing risks in our statistical model. Following Statistics Denmark classification system (Statistics Denmark, 2016), the exit or the end of employment is identified by not working for two consecutive months. We compute employment duration as the number of weeks from the first week of 2009 (aged 59) until an exit takes place. The exit route is determined by the post-employment status. However, if the status is retirement and the individual starts receiving retirement pension before employment terminates, we define entrance into the retirement program as the end of the employment period.

Because our data cover the whole period from 2009 to 2016 for all individuals in the sample, censoring only

occurs at the last week in 2016 (end of observation period) when no exit has taken place. Table 2 reports the number and share of observed transitions into the two risks in the sample. We can see that 86% of the sample enter a retirement program. Among those individuals, 3,272 (40.01%) enter the old age pension, 3,755 (45.92%) enter the early retirement pension, and only 16 (0.20%) enter the disability pension. The low occurrence of disability pension is quite reasonable due to the construction of our sample. More details on the construction of the duration and the determination of exit route are provided in Supplementary Material S.2.

Table 2 Number and Share of Transitions into Risks

Risk	Number	Share (%)
1 Retirement	7,043	86.12
2 Others	944	11.54
Right-censored	191	2.34
Total	8,178	100.00

By using weekly information for a period of 8 years, we have (nearly) continuous duration data. Figure 1 Panel A shows the histogram of durations with exit to retirement and durations for right-censored observations. The minimum duration is 2 weeks, corresponding to the second week in 2009. The maximum duration is 418 weeks, corresponding to the last week in 2016. The highest frequencies are for durations from 266 to 315 weeks, which correspond to year 2014 when individuals in the sample turn 65 and satisfy the age requirement for old age pension. We can see that those who exit to old age pension all do so in 2014. Compared with old age pensioners, the distribution of duration for early retirement pensioners is sparser. The first peak corresponds to year 2009 when the individuals turn 60 and satisfy the age requirement for early retirement pension. The second peak at around 130 weeks corresponds to year 2011 when the individuals turn 62, which is the shortest time (two years) for early retirement pension deferral. The longest duration, 418 weeks, corresponds to censored observations. They are those who defer old age pension from 2014 when they satisfy the age requirement until at least the end of 2016, suggesting that most individuals defer old age pension for either a long or a short period as the density in between is zero.

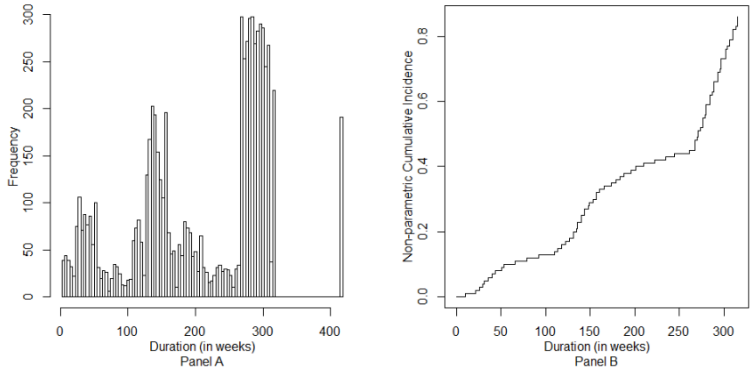

Fig.1 Descriptive Statistics of the Distribution of Employment Durations

Figure 1 Panel B shows the non-parametric estimates of the cumulative incidence curve for exit to retirement. This is the implied curve from the competing risks quantile regression model without regressors. The three sharp increases in Panel B correspond to the three peaks in Panel A. The estimated cumulative incidence almost doubles for duration from 266 weeks to 315 weeks, suggesting that almost half of the individuals in the sample

retire at these durations. The curve in Panel B cannot be obtained for durations longer than 315 weeks because the estimated cumulative incidence reaches its plateau level due to lack of observed transitions, which makes it non-invertible and the conditional quantile goes to infinity (see Section 5).

Our data come from 16 linked administrative registers and contain information on various personal, household, and firm characteristics, including demographics, education, income, pension, employment, socioeconomic status, health, criminal records and a wealth of company (employer) statistics. Figure 2 gives an overview of the registers. The second column shows the information that the register contains. The third column shows the names of the registers. The last column gives some examples of the variables within each register and shows the number of within-register variables, denoted by A. Each register forms a group of 2 to 58 variables. We provide a detailed description of the data source in Supplementary Material S.1. The construction of the estimation sample is described in Supplementary Material S.3. Descriptive statistics for all variables are shown in Table S.1 and documentation of selected variables is given in Table S.2 in Supplementary Material S.4.

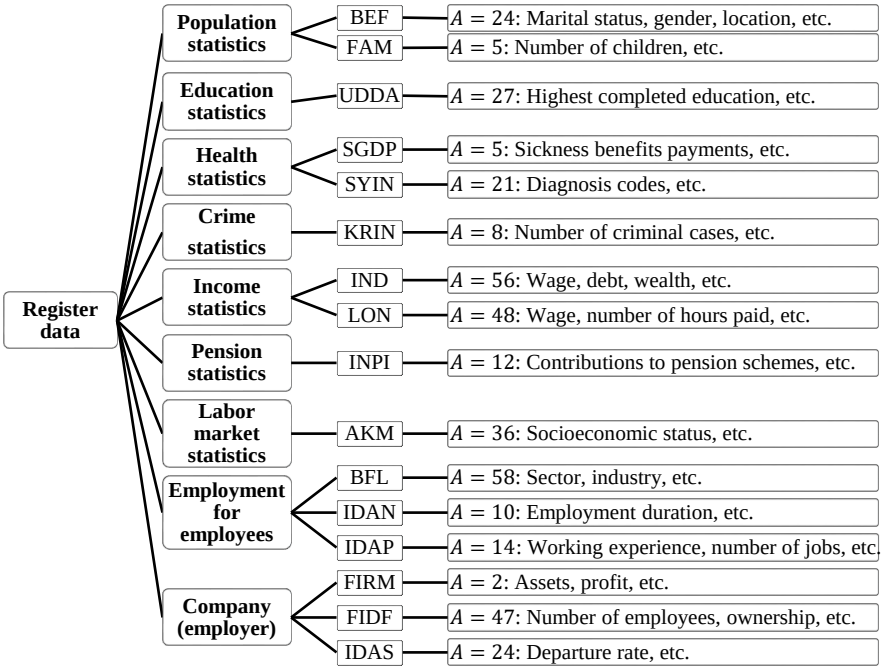

Fig.2 Overview of the Structure of the Linked Register Data

5 Methodology

Our analysis applies the competing risks quantile regression framework by Peng and Fine (2009) to analyze exits routes out of the labor force. Our explained variable Y is employment duration as worked out in Section 4. Let the conditional distribution of Y be $F_Y(y|X) = Pr(Y \leq y|X)$ and X contains K regressors. The τ'th conditional quantile of $F_Y(y|X)$ is $Q_Y(\tau|X) = F_Y^{-1}(\tau|X) = inf\{y: F_Y(y|X) \geq \tau\}$. Koenker and Bassett (1978) consider a linear representation of the conditional quantile $Q_Y(\tau|X) = X\beta(\tau)$, where $\beta(\tau)$ is a vector of unknown parameters with length K. The estimator $\hat{\beta}(\tau)$ can be obtained by minimizing $\sum_{i=1}^{N}\left(\tau -\right.$

$\mathbb{1}_{\{y_i \leq x_i' b(\tau)\}}\big)(y_i - x_i' b(\tau)) \stackrel{\text{def}}{=} \sum_{i=1}^{N} \rho_\tau(y_i - x_i' b(\tau))$ with respect to $b(\tau)$, where $\rho_\tau(u) \stackrel{\text{def}}{=} (\tau - \mathbb{1}_{\{u \leq 0\}})u$ is known as the check function and $\mathbb{1}(\cdot)$ is an indicator function.

Competing risks model refers to duration analysis with several potential failure types, or risks. In our application we consider two risks $r = 1,2$, which are retirement and other exits routes as described in Section 4. We also need to accommodate that our duration data are censored at the end of the observation period. Let T_r and C denote event time and an independent censoring point respectively. Let the minimum duration be $U = min_r\{T_r\}$ and the corresponding risk be $\epsilon = arg\, min_r\{T_r\}$. The observed duration is $T = min(U, C)$. The observed failure type is $\Delta = \mathbb{1}_{\{U \leq C\}}\epsilon$. The competing risks model is plagued by non-identifability issues (Cox, 1962), because only the shortest of the competing risks duration can be observed and the dependence structure between competing risks is unknown. The identification bounds for parameters of interest are typically wide (Peterson, 1976). By imposing additional restrictions, Heckman and Honoré (1989) show identifiability, but restrictions are difficult to verify in practice (Femanian, 2003). Instead of focusing on the distribution of the competing random variables, it is more practical to consider cause specific distributions or subdistributions, which describe observable transition patterns. For a link between these approaches, see Emura et al. (2019). Various models for cumulative incidences and subdistributions have been suggested, with one of the most popular being the Cox type semiparametric proportional hazards model by Fine and Gray (1999). It is important to mention that cumulative incidences and subdistributions are identifiable without knowledge of the dependence structure between competing risks. Peng and Fine (2009) suggest a competing risks quantile regression model for the cumulative incidence, where the latter is $F_r(t|X) = Pr(T_r \leq t, \Delta = r|X)$ for risk $r = 1,2$. Dlugosz et al. (2017) elaborate that the implied restrictions of partial regressor effects on cumulative incidences are less restricted in the quantile regression model than in proportional hazards models (e.g. Fine and Gray, 1999). Given that we expect in our empirical analysis a variety of effects on transitions at different durations due to the different retirement programs, we choose the competing risks quantile regression model to permit for sufficient flexibility.

The focus in our analysis is on risk 1, the transition into retirement. Risk 2, in contrast, is a pooled exit state that contains everything else than retirement and therefore does not have a clear interpretation. Results for risk 1 do not depend on whether the other risks are pooled or not, because we consider cumulative incidences. The τ'th conditional quantile of the cumulative incidence for risk 1, $F_1(t|X)$, is $Q_1(\tau|X) = inf\{t: F_1(t|X) \geq \tau\}$. Assume $Q_1(\tau|X) = g\big(X\beta(\tau)\big)$, where $g(\cdot)$ is a known monotone link function and $0 < \tau_L \leq \tau \leq \tau_U < 1$. Q_1 is therefore nonlinear in $\beta(\tau)$ but the direction of the partial effect of one regressor on the conditional quantile is determined by the sign of the relevant parameter. A positive parameter increases the conditional quantile, which corresponds to later transitions times into retirement. Moreover, the parameters are directly informative about the relative size of the effect of a regressor compared to other regressors. Looking at the parameters is therefore informative in this model, despite its complexity. τ_U is less than one because of the nature of competing risks models. It reflects that the share of observable transitions is less than one for each risk. The cumulative incidence has a plateau level and the conditional quantile does not exist above this level. This is not a disadvantage of modelling conditional quantile functions as they can still attain large values and explode at τ_U. The model is therefore capable of producing results for long durations even if τ_U is considerably below one. The sample analogue of (T, Δ, C, X) is denoted as $(t_i, \delta_i, c_i, x_i)$. In the case of no censoring, similar to the linear quantile regression model, the estimator $\hat{\beta}(\tau)$ can be obtained by minimizing $\sum_{i=1}^{N} \rho_\tau(g^{-1}(t_i^*) - x_i' b(\tau))$ with respect to $b(\tau)$, where x_i is a $K \times 1$ vector of regressors and $t_i^* =$

$\mathbb{1}_{\{\delta_i=1\}}t_i + \mathbb{1}_{\{\delta_i\neq 1\}} \times \infty$, which is equivalent to solving equation $N^{-\frac{1}{2}}\sum_{i=1}^{N} x_i'\left(\mathbb{1}_{\{g^{-1}(t_i)\leq x_i'b(\tau),\delta_i=1\}} - \tau\right) = 0$. In the case of independent censoring, $\hat{\beta}(\tau)$ is the solution to $S_N(b(\tau),\tau) = 0$, where

$$S_N(b(\tau),\tau) = N^{-\frac{1}{2}}\sum_{i=1}^{N} x_i'\left(\frac{\mathbb{1}_{\{g^{-1}(t_i)\leq x_i'b(\tau),\delta_i=1\}}}{\hat{G}(t_i)} - \tau\right),$$

and $\hat{G}(\cdot)$ is the Kaplan-Meier estimator for $Pr(C \geq T|X)$. Because $S_N(b(\tau),\tau) = 0$ may not have an exact solution due to noncontinuity, Peng and Fine (2009) define a generalized solution and show that it is equivalent to minimizing the following ℓ_1-type convex function

$$\begin{aligned} U_N(b(\tau),\tau) = &\sum_{i=1}^{N} \mathbb{1}_{\{\delta_i=1\}}\left|\frac{g^{-1}(t_i) - x_i'b(\tau)}{\hat{G}(t_i)}\right| \\ &+ \left|M - b(\tau)'\sum_{i=1}^{N}\frac{-x_i\mathbb{1}_{\{\delta_i=1\}}}{\hat{G}(t_i)}\right|, \\ &+ \left|M - b(\tau)'\sum_{i=1}^{N} 2x_i\tau\right| \end{aligned}$$

where M is a very large positive number. They prove consistency and asymptotic normality of $\hat{\beta}(\tau)$ under some regularity conditions. The former requires four regularity conditions, C1-C4 in Peng and Fine (2009). C1 is a standard assumption on the censoring. In our application, the censoring is due to the end of the observation period and is compatible with their restriction. C2 and C3 are two technical restrictions that are common for censored quantile regression: C2 is uniform boundedness of covariates, which also holds with our data. C3 requires the QR coefficients to vary smoothly in τ and the derivative of the cumulative incidence to be bounded from above. While it is difficult to anticipate the former, the latter essentially rules out mass points that lead to jumps. This restriction is possibly violated with our data, because the descriptive analysis has shown that there are mass points in the distribution of observed durations (Figure 1 Panel A). Condition C4 is that the cumulative incidence reaches its plateau level at τ_U and rules out flat intervals before the final plateau level is reached. Given the evidence provided in Figure 1, this condition could be also violated, as the density between the mass points is quite low. In order to check whether our results are sensitive to the presence of mass points, we adopt the idea of Machado and Silva (2005) and smooth the distribution. This is done by adding an independent random noise to the discrete mass points and thus the distribution becomes continuous. Machado and Silva show that the conditional quantiles of the constructed variable have a one-to-one relationship with those of the original variable, and the conditional quantiles of the original variable can be consistently estimated under mild assumptions. We follow their implementation, use uniform and truncated normal distribution for the independent random noise and adapt the parameters to different intervals between mass points. By doing so, we find general robustness of our results. Moreover, τ_U does not change when employing their methods. These checks therefore do not provide evidence for our point estimates being adversely affected by the possible violations.

Besides point estimates of coefficients and variance estimators, Peng and Fine (2009) suggest a trimmed mean statistic to summarize the effect over quantiles. The trimmed mean effect estimator is defined as $\int_{\tau_L}^{\tau_U}\hat{\beta}(\tau)d\tau/(\tau_U - \tau_L)$. It measures the mean of the estimated effect of a regressor on the conditional quantiles of cumulative incidence curve from τ_L to τ_U and thus can act as a summary statistic for the average effect of regressors over quantiles. In practice, we use Riemann sum to approximate the integral. They also suggest a Wald-type constant test on whether a regressor has a constant effect on the cumulative incidence quantiles. For the constant test, the null hypothesis is $H_0: \beta(\tau) = \rho_0, \tau \in [\tau_L, \tau_U]$, where ρ_0 is an unspecified constant, and the test statistic is derived on the grounds of the trimmed mean effect estimator.

In our analysis we use the exponential function as the link function $g(\cdot)$. The model is estimated for $\tau \in [0.01, \tau_U]$ with a step size of 0.01, where τ_U is determined automatically as a value that corresponds to an cumulative incidence that is lower than its plateau value, resulting from condition C4 of Peng and Fine (2009). Statistics are computed in R 3.4.2 (R Core Team, 2017) using the cmprskQR (v0.9.2; Dlugosz et al., 2019) and the quantreg (v5.33; Koenker, 2017) packages. As it will be shown in Section 6, we obtain different τ_U when different regressors are included in the model, the reason is that violations of the regularity conditions occur at different quantiles for different regressors. In general, we find that the lower the number of regressors, the higher the τ_U. As there are also some issues with the inference at the higher quantiles, we mainly focus on reporting the results for $\tau_U = 0.58$ in the main text. Additional results along with additional robustness checks are presented in Supplementary Materials S.5-S.8.

Given the high dimensionality of our regressor matrix, it is natural to apply variable selection methods for regularization due to its superiority over traditional sequential elimination methods.

Due to increases in computing power and progress in methodology, penalized regression and shrinkage methods become increasingly developed and popular among practitioners in a wide range of statistical applications. The idea is to minimize a penalized objective function, where a penalty, $P_\lambda(\beta(\tau), \tau)$, is added to the original objective function. The penalized objective function depends on the parameters and an additional non-negative tuning parameter, or the regularization parameter λ. A number of penalizations have been suggested, including ridge regression (Hoerl, 1962), the least absolute selection and shrinkage operator (Lasso) by Tibshirani (1996), ℓ_1-penalized linear quantile regression (Belloni and Chernozhukov, 2011) and the adaptive Lasso (Zou, 2006). The (adaptive) Lasso estimator can select relevant individual variables, but does not perform well when variables have group structure, such as dummy variables formed from a categorical variable, or in our case, variables within one register. Rather than identifying relevant individual variables, sometimes the objective is to identify relevant variable groups and set the coefficients of all variables in the irrelevant groups to zero. Unfortunately, the (adaptive) Lasso also selects variables of irrelevant groups in this case. Yuan and Lin (2006) extend the Lasso to group variable selection and introduce the group Lasso, which either selects or drops all variables of a group. However, in many cases, only some variables within each group are relevant for the outcome and we want to include only those relevant individual variables in the analysis. In order to select relevant individual variables within groups, Huang et al. (2009) introduce the group bridge method and further extend the Lasso to bi-level selection. The penalty term of group bridge is a non-convex bridge penalty (Fu, 1998) for groups and a ℓ_1-type penalty for within-group variables. This method can select both relevant groups and relevant individual variables within those groups. Huang et al. (2009) prove the group selection consistency, but do not prove selection consistency for within-group individual variables. Due to the ℓ_1-type penalty, the group bridge shares similar shortcomings of the Lasso. Similar to the change from the standard Lasso to the adaptive Lasso, the adaptive group bridge modifies the ℓ_1-type penalty to a weighted ℓ_1-type penalty. See Huang et al. (2012) for a survey on group selection and bi-level selection methods.

Ahn and Kim (2018) study the behavior of (adaptive) group bridge applied to competing risks quantile regression. Suppose that the K explanatory variables belong to J groups. In each group there are A_j explanatory variables denoted by β_{jk} where $j = 1, \dots, J$ and $k = 1, \dots, A_j$. The objective function of the penalized competing risks quantile regression with (adaptive) group bridge penalty is as follows,

$$W_N(b(\tau), \tau) = U_N(b(\tau), \tau) + \lambda_N \sum_{j=1}^{J} A_j^{1-\gamma} \left(\sum_{k=1}^{A_j} \left({|b_{jk}(\tau)|} \Big/ {|\tilde{\beta}_{jk}(\tau)|^{\nu}} \right) \right)^{\gamma}.$$

With the ℓ_1-type penalty in the second term, an individual variable can be selected or dropped according to

the effects from both itself and its group. The (adaptive) group bridge penalty has four tuning parameters $(\lambda_N, \gamma, \nu, \tilde{\beta})$. γ is the tuning parameter for the bridge penalty that is between zero and one. $A_j^{1-\gamma}$ is the group level weight to adjust for sizes of groups. $\tilde{\beta}_{jk}$ is an initial consistent estimator for β_{jk}. ν is the non-negative individual level weight parameter for the k^{th} variable within group j. There are some special cases for different values of the tuning parameters. When $\nu = 0$, the adaptive group bridge becomes group bridge, as the variables within each group are treated with the same individual level weight, which is one. When $\gamma = 1$ and there is no group structure, the adaptive group bridge becomes a simple $\ell 1$-type penalized estimator, as the variables are not penalized based on the groups that they belong to. Specifically, it reduces to the Lasso when $\nu = 0$, and the adaptive Lasso when $\nu > 0$.

Minimization of $W_N(b(\tau), \tau)$ itself is not easy due to the non-convexity of this function. Similar to Huang et al. (2009), Ahn and Kim (2018) propose that through variable augmentation, minimizing $W_N(b(\tau), \tau)$ with respect to $b(\tau)$ is equivalent to minimizing $\widetilde{W}_N(b(\tau), \theta, \tau)$ with respect to $(b(\tau), \theta)$,

$$\widetilde{W}_N(b(\tau), \theta, \tau) = U_N(b(\tau), \tau) + \xi_N \sum_{j=1}^{J} \left(\left(\frac{\theta_j}{A_j} \right)^{1-\frac{1}{\gamma}} \sum_{k=1}^{A_j} \left({|b_{jk}(\tau)|} \Big/ {|\tilde{\beta}_{jk}(\tau)|^{\nu}} \right) \right) + \xi_N \sum_{j=1}^{J} \theta_j$$

$$\theta_j = A_j^{1-\gamma} \left(\frac{1-\gamma}{\gamma} \right)^{\gamma} \left(\sum_{k=1}^{A_j} \left({|\beta_{jk}(\tau)|} \Big/ {|\tilde{\beta}_{jk}(\tau)|^{\nu}} \right) \right)^{\gamma},$$

where the tuning parameter ξ_N is a reparameterization of λ_N. They prove that under some conditions, the group bridge selects group variables consistently; the adaptive group bridge not only selects group variables consistently, but also selects within-group individual variables consistently, and thus possesses the oracle property. In addition to conditions C1-C4 by Peng and Fine discussed above, it is required that the number of groups, the number of variables in groups and the magnitude of parameters are restricted as the number of observations goes to infinity. Given that we have a fixed number of variables and groups that are much smaller than the number of observations, we do not consider these restrictions as crucial in our application.

There are different ways to set the values of the tuning parameters $(\xi_N, \gamma, \nu, \tilde{\beta})$. Ahn and Kim (2018) set $(\gamma, \nu) = (1/2, 1)$. They use two initial estimators for $\tilde{\beta}$ – one is the group bridge estimator, and the other is the competing risks quantile regression estimator. Overall, they apply three methods – group bridge, adaptive group bridge with group bridge as an initial estimator, adaptive group bridge with competing risks quantile regression as an initial estimator. They use a BIC-type criterion to select the best tuning parameter ξ_N for the three methods. Fu (1998) chooses γ among 40 equidistant values in $[1,3]$ and selects (λ, γ) using a two-dimensional generalized cross-validation for bridge regression. Zou (2006) chooses ν from $\{0.5,1,2\}$ and selects (λ, ν) using a two-dimensional cross-validation for adaptive Lasso. He also mentions that one can treat the initial estimator $\tilde{\beta}$ as the third tuning parameter and perform a three-dimensional cross-validation to find an optimal triple $(\gamma, \nu, \tilde{\beta})$. Motivated by the literature, we propose a four-dimensional selection procedure. Define a 4-dimensional grid $\mathcal{G} = \xi_N \times \gamma \times \nu \times \tilde{\beta}$ for the tuning parameters. We use the BIC-type selection criterion proposed by Ahn and Kim (2018),

$$\frac{2}{N} U_N(\hat{\beta}(\tau), \tau) + p_N \, ln(K) \frac{ln(N)}{2N},$$

where K is the number of explanatory variables, N is the number of observations, and p_N is the number of nonzero coefficients, i.e. the selected variables, to model degrees of freedom. We choose γ from $\{0.25,0.5,0.75\}$, ν from $\{0,0.5,1,1.5,2\}$, and $\tilde{\beta}$ from group bridge estimator and competing risks quantile regression estimator. Due to the two choices of initial estimators for the adaptive group bridge, we have three

methods in total. We select the optimal tuning parameters $(\xi_N^*, \gamma^*, \nu^*, \tilde{\beta}^*)$ that lead to the smallest BIC-type selection criterion value. First we set $(\gamma, \nu, \tilde{\beta})$ to a certain combination of the parameter values and find the optimal regularization parameter ξ_N, and then we continue this process for another combination of $(\gamma, \nu, \tilde{\beta})$ until we find the optimal $(\xi_N^*, \gamma^*, \nu^*, \tilde{\beta}^*)$. Besides the optimal tuning parameters among all three methods, we also obtain the optimal tuning parameters for each method. The complete algorithm for the (adaptive) group bridge is given in Appendix 1.

Inference for sparse estimators, such as Lasso type penalized regression, is still a rapidly developing area. These models are characterized by slower than $\sqrt{N}$ rate of convergence, which can lead to sizable finite sample biases (Chernozhukov et al., 2018), in particular for $K>N$. In order to derive the asymptotic distribution of sparse estimators, it is typically assumed that only the correct variables have been selected by the algorithm. Usual bootstrap procedures may be invalid, because the asymptotic distribution of the estimators can be non-smooth due to the selection process. For a discussion of these issues and valid inference procedures for sparse models see for example Van de Geer et al. (2014) and Zhang and Zhang (2014). Alternative approaches include sample splitting (Meinshausen et al., 2009), covariance test (Lockhart et al., 2014), exact post-selection inference (Lee et al., 2016), etc. See Taylor and Tibshirani (2015) for a survey on post-selection inference, or selective inference. One naive inference method is to use unpenalized regression models with the selected variables. This is known as the OLS post-Lasso estimator for the linear mean regression model (Belloni and Chernozhukov, 2013, Hastie et al., 2015: 301). Inference on the grounds of this approach is too optimistic because it ignores the uncertainty of the selection process and is only valid under strong assumptions. Modifications of the residual bootstrap have been developed for the penalized linear mean regression model (Chatterjee and Lahiri, 2011, 2013) and the wild residual bootstrap has been shown to be valid for penalized linear quantile regression (Wang et al., 2018). In econometrics, double machine learning has been shown to reduce finite sample bias and correct the distribution of the estimator (Chernozhukov et al., 2018). Chen and Tien (2019) suggest double machine learning for linear instrumental variable quantile regression and find in simulations that it gives efficiency of the estimate that is close to a model with known true regressor set. While all these approaches are promising, they have not yet been adapted to competing risks quantile regression for the cumulative incidence and therefore cannot be directly applied in our analysis. We therefore report post-Lasso inference statistics as in Ahn and Kim (2018), where we consider asymptotic and nonparametric bootstrap statistics. Moreover, we report full model selection nonparametric bootstrap inference, although the latter only for one method and one quantile due to being computationally too demanding. It is well known that asymptotic inference for the quantile regression model can be sensitive to finite sample errors in estimated conditional error distributions. This is potentially an issue in our application due to intervals with low density of dependent variable (compare Figure 1). We report bootstrap statistics as a robustness check. Finite sample biases of sparse estimators should be less relevant the larger the sample and the smaller the number of regressors relative to the sample size. Given that our estimation sample has more than 8000 observations with "only" several hundred variables, we do not expect our results to contain large finite sample biases.

6 Results

In the following, we focus on the estimation results for three quantiles: 0.11, 0.25, and 0.31. For each of these three quantile, we select the best tuning parameters (ξ_N, γ, ν) according to the BIC-type selection criterion for three methods: group bridge, adaptive group bridge with group bridge as the initial estimator, and adaptive

group bridge with competing risks quantile regression as the initial estimator. We report selection results with estimated coefficients of both the penalized methods and the post-Lasso type unpenalized competing risks quantile regression model at the three quantiles in Table 3. For inference, we report nonparametric bootstrap p-value for the post-Lasso type estimation. Only variables selected by at least one method are shown in the table for each quantile. We provide a comparison of the asymptotic standard error and nonparametric bootstrap standard error for the post-Lasso type estimation, and the full model selection nonparametric bootstrap standard error for the non-post penalized methods, and discuss the empirically observed limitation of the post-Lasso type inference in presence of multicollinearity in Supplementary Material S.5.

For $\tau = 0.11$, each method selects 3 to 4 registers and around 10 within-register variables. For $\tau = 0.25$, each method selects 3 registers and around 10 within-register variables. For $\tau = 0.31$, each method selects 1 to 2 registers and around 4 within-register variables. So, the number of selected variables decreases with τ, and all methods reduce model dimension considerably – from 16 register to less than 5 registers, and from 397 variables to less than 15 variables. For $\tau = 0.11$, penalized methods are more likely to shrink the magnitude of the estimates compared to unpenalized estimation; however, this relationship is reversed for the other quantiles. Regarding model size, group bridge tends to select slightly more variables than adaptive group bridge. The three methods agree on most selected registers and within-register variables, and the signs of the estimates. Most selected variables are significant according to the post-Lasso type bootstrap inference.

Multicollinearity causes some problems for the selection and inference. Some registers and variables are selected by only one method. It is then less convincing that these registers and variables are important. The significance levels of the financial variables in the income register IND are considerably lower than variables in other registers. In addition, post-Lasso estimation and penalized methods give different signs to the estimated coefficients of two variables in IND. We will explain these findings in relation to high multicollinearity later.

Of the 16 registers, the education, health, and crime registers are not selected. Compared with Figure 2, we are only left with registers covering labor market, employment, population and financial statistics. This may partly contradict the selection results in previous studies, but it can be explained to some extent. First, education has ambiguous effects on retirement (Kallestrup-Lamb et al., 2016). Second, the sample of this study consists of rather healthy individuals and from the data we know that only few people have criminal records, which could explain the omission of health and crime information. For the definitions of selected within-register variables, we again refer to Table S.2 in Supplementary Material S.4 for links to the detailed documentation of selected variables on the relevant websites of Statistics Denmark.

For the labor market register AKM, 4 occupation variables, 2 industry variables, and 2 socioeconomic status variables are selected. The three methods agree on most selected variables. Except for *Industry: Energy supply*, *Occupation: Professional* and *Employed: High level*, the other selected occupations and industries all have negative signs, suggesting shorter employment duration with exit to retirement and higher transition probability from employment to retirement at the corresponding quantile. Among them, *Occupation: Professional* and *Employed: High level* indicate being employed in work that requires knowledge at the highest level in the respective fields. These two dummy variables are highly collinear – out of the 8,178 observations, only 9 values are different. So we can see that at the 0.25 quantile, group bridge and adaptive group bridge with competing risks quantile regression as the initial estimator select *Occupation: Professional*, while adaptive group bridge with group bridge as the initial estimator selects *Employed: High level*. However, it seems that most industries and occupations are unimportant for transitions into retirement, as there are 9 occupations and 18 industries before the selection. Employment history variables such as work experience and unemployment

Table 3 Estimation Results for Employment Duration With Exit to Retirement

	Register	Within-Register Variable	GB			AGB-GB			AGB-CRQR		
			Non-Post	Post		Non-Post	Post		Non-Post	Post	
			Est.	Est.	P-Value	Est.	Est.	P-Value	Est.	Est.	P-Value
0.11 Quantile	AKM	Occupation: Operation/Transport				-0.178	-0.383	0.016			
		Occupation: Manual	-0.243	-0.573	0.000	-0.619	-0.742	0.000	-0.522	-0.607	0.000
		Occupation: Military	-0.542	-1.038	0.000	-1.032	-1.149	0.080			
		Occupation: Professional	0.261	0.240	0.016	0.253	0.236	0.032	0.239	0.234	0.004
		Municipal employment	-0.225	-0.294	0.000						
		Industry: Healthcare							-0.216	-0.226	0.000
		Employed: Basic level	-0.207	-0.369	0.000	-0.448	-0.414	0.000	-0.469	-0.535	0.000
	BEF	Date of birth	0.006	0.011	0.000						
		Divorced				0.290	0.449	0.004			
		Family reference person				-0.369	-0.456	0.000			
		Male	0.151	0.316	0.000						
	BFL	Workplace sector: Municipally owned institution[a]				-0.333	-0.464	0.000			
		Workplace location: Copenhagen City				0.064	0.171	0.056			
	FIDF	Workplace sector: Municipally owned institution[b]							-0.208	-0.363	0.000
	IDAP	Insured				-0.734	-0.697	0.000			
	IND	AM-income (mil.)	0.692	0.684	0.000				0.985	0.665	0.000
		Other capital income (mil.)	0.710	0.769	0.220						
		Contributions to union, unemp. insurance, and PEW (thous.)	-0.013	-0.022	0.016						
		ATP contributions (thous.)							0.078	0.088	0.004
		Contributions to PEW (thous.)	-0.176	-0.143	0.000				-0.152	-0.185	0.000
		Bonded debt (mil.)	0.028	0.058	0.372						
		Pension income (mil.)							0.788	0.567	0.604
0.25 Quantile	AKM	Occupation: Operation/Transport				-0.183	-0.229	0.128			
		Occupation: Manual	-0.178	-0.067	0.032	-0.571	-0.269	0.000	-0.214	-0.061	0.012
		Occupation: Military	-0.485	-1.232	0.000	-1.163	-1.346	0.020			
		Occupation: Professional	0.233	0.025	0.032				0.340	0.014	0.172
		Municipal employment	-0.204	-0.026	0.016	-0.293	-0.085	0.008	-0.199	-0.014	0.096

(Continued)

Table 3 Estimation Results for Employment Duration With Exit to Retirement (Continued)

	Register	Within-Register Variable	GB			AGB-GB			AGB-CRQR		
			Non-Post	Post		Non-Post	Post		Non-Post	Post	
			Est.	Est.	P-Value	Est.	Est.	P-Value	Est.	Est.	P-Value
0.25 Quantile	AKM	Industry: Energy supply				0.281	0.144	0.648			
		Industry: Healthcare							-0.036	-0.023	0.020
		Employed: High level				0.327	0.074	0.104			
		Employed: Basic level	-0.181	-0.022	0.076	-0.365	-0.126	0.000	-0.258	-0.023	0.020
	BEF	Date of birth	0.009	0.006	0.000				0.009	0.006	0.000
		Divorced				0.326	0.146	0.008			
		Family reference person				-0.365	-0.126	0.008	-0.040	-0.012	0.108
		Male	0.119	0.020	0.036						
	IDAP	Insured				-0.914	-0.622	0.000			
	IND	AM-income (mil.)	0.711	0.244	0.228				0.900	0.317	0.004
		Other capital income (mil.)	0.538	-0.100	0.560						
		Contributions to union, unemp. insurance, and PEW (thous.)	-0.015	-0.003	0.084						
		Contributions to PEW (thous.)	-0.197	-0.134	0.000				-0.203	-0.143	0.000
		Total capital income (mil.)							0.436	0.163	0.064
		Bonded debt (mil.)	0.033	0.027	0.580						
		Taxable personal income (mil.)	0.059	0.069	0.736						
		Interest expense on mortgage debt (mil.)	0.273	0.671	0.180						
0.31 Quantile	BEF	Date of birth	0.036	0.006	0.000	0.025	0.006	0.000	0.036	0.006	0.000
	IND	AM-income (mil.)	0.916	0.722	0.000				0.965	1.293	0.004
		Contributions to PEW (thous.)	-0.210	-0.128	0.000				-0.176	-0.133	0.000
		Rental value of own housing (mil.)							0.270	0.636	0.000
		Salary income (mil.)							0.147	-0.581	0.032
		Bonded debt (mil.)	0.092	0.092	0.016						

Note: GB refers to the group bridge estimator. AGB-GB refers to the adaptive group bridge using group bridge as the initial estimator. AGB-CRQR refers to the adaptive group bridge using competing risks quantile regression as the initial estimator. Non-Post refers to estimates from the penalized methods. Post refers to estimates from the post-Lasso type unpenalized competing risks quantile regression using selected variables from the respective penalized methods. We use nonparametric bootstrap p-values. Amounts in DKK.

experience are also unselected, possibly because the sample consists of employees with rather similar work experience and little unemployment experience.

For the population register BEF, 1 marital status variable, 2 gender related variables, and 1 age related variable are selected. Since the sample consists of individuals born in 1949, they only differ in age in terms of the number of weeks from 1 Jan to the date when one is born in the given year. We compute *Date of birth* as the number of weeks from 1 Jan to the date when one is born in order to capture the effect of age-related eligibility on entering a pension program. It is selected by almost all three methods and is significant at all three quantiles. Later we will show that *Date of birth* is the only variable that is selected for the higher quantiles. The sign is always positive, suggesting that the employment duration with exit to retirement is very sensitive to and positively affected by the eligibility of entering a pension program. The family reference person is taken to be the women in a heterosexual couple family and the oldest person in other families. So by definition, *Male* and *Family reference person* are highly negatively correlated dummy variables – indeed, more than 90% of the observations have opposite values. We can see that group bridge chooses *Male* and two adaptive group bridge methods choose *Family reference person* at the 0.11 and 0.25 quantile, and they have opposite signs as well. These two gender related variables and *Divorced* become unselected at the 0.31 quantile, suggesting that male and divorced individual is more likely to have an observed transition into retirement at lower quantiles.

Some registers are selected by only one method. The dummy variables *Municipal employment*, *Workplace sector: Municipally owned institution*[a], and *Workplace sector: Municipally owned institution*[b] come from different registers – AKM, BFL and FIDF respectively. However, they are highly correlated – more than 90% of the observations have the same values. The three methods select one of the three variables respectively at the 0.11 quantile. Only adaptive group bridge with group bridge as the initial estimator selects the IDAP register. The selected within-register variable *Insured* indicates whether one is insured, including part-time and full-time insured. The pairwise correlation among *Insured*, *Contributions to union, unemployment insurance, and PEW*, and *Contributions to PEW* are around 0.6, the latter two of which belong to the register IND. Similar to the variable that indicates workplace sector, the other two methods select the variables in the income register IND instead of *Insured*. Although these methods give different opinions on the relevant registers and within-register variables, the selection can indicate what kind of information is important to some extent, because the selected variables are highly correlated and share similar information.

For the income register IND, income related and debt related variables all have positive signs for all quantiles, suggesting that people with higher income and more debt tend to work longer. This finding is in line with Kallestrup-Lamb et al. (2016), who find that the effect of income variables on the retirement decision is dominated by substitution effect in the tradeoff between leisure and income. *Contributions to PEW* is a strong indicator of observed transition from employment into retirement, as this variable has strong and negative effects for all three quantiles. Multicollinearity is more prevalent for financial variables. Only group bridge and adaptive group bridge with competing risks quantile regression as the initial estimator select register IND for all three quantiles. The pairwise correlations among *Other capital income*, *Rental value of own housing*, *Bonded debt*, and *Interest expense on mortgage debt* are from 0.4 to 0.95. The pairwise correlation between *Other capital income* and *Total capital income* is around 0.92. The pairwise correlations among *AM-income*, *Salary income* and *Taxable personal income* are around 0.95. Group bridge selects two highly correlated variables at the 0.11 quantile, and more at the 0.25 quantile. The significance levels of these variables are lower than when only one of them is selected. Adaptive group

bridge using competing risks quantile regression as the initial estimator selects two highly correlated variables at the 0.31 quantile. The post-Lasso type estimates have opposite signs compared with the penalized methods for two variables – *Other capital income* of group bridge at the 0.25 quantile, and *Salary income* of adaptive group bridge using competing risks quantile regression as the initial estimator at the 0.31 quantile.

In all, we find that all three methods perform well for registers except IND in the sense that they select only one of the highly collinear variables and the post-Lasso type estimation agrees with the penalized methods on the signs of the estimates. For the income register IND, group bridge suffers more from multicollinearity compared with adaptive group bridge. However, only the within-register variable selection performance is affected, as the group bridge and adaptive group bridge with competing risks quantile regression as the initial estimator both select this register. We can also compare the methods through the BIC-type criterion value. Table 4 shows the BIC-type selection criterion value for all combinations of $(\xi_N, \gamma, \nu, \tilde{\beta})$. We find that they do not have too large differences in terms of the criterion value. Adaptive group bridge with group bridge as the initial estimator performs worse than the other two methods for all three quantiles. Group bridge performs slightly better than adaptive group bridge with competing risks quantile regression as the initial estimator for the 0.11 and 0.31 quantile, and the relationship is reversed for the 0.25 quantile, showing that group bridge performs better when it is not largely affected by multicollinearity.

Table 4 BIC-Type Selection Criterion Value for Various Models

			GB	AGB-GB				AGB-CRQR				LASSO	
			Tuning Parameter ν										
			BIC-Type Selection Criterion										CV
			0	0.5	1	1.5	2	0.5	1	1.5	2	0	0
Tuning Parameter γ	0.11 Quantile	0.25	15.5813	15.6136	15.6078	15.6065	15.6049	15.5797	15.5816	15.5805	15.5824		
		0.5	15.5811	15.6108	15.6047	15.6040	15.6061	15.5829	15.5813	15.5819	15.5842		
		0.75	**15.5767**	**15.5984**	15.6033	15.6069	15.6137	15.5804	15.5811	**15.5794**	15.5852		
		1										**15.5997**	15.8598
	0.25 Quantile	0.25	13.2559	13.2735	13.2827	13.2772	13.2774	13.2476	13.2486	13.2467	13.2461		
		0.5	**13.2503**	13.2727	**13.2712**	13.2784	13.2766	13.2452	13.2478	13.2481	13.2467		
		0.75	13.2511	13.2748	13.2766	13.2793	13.2905	**13.2431**	13.2457	13.2460	13.2491		
		1										**13.2621**	13.5748
	0.31 Quantile	0.25	13.0423	13.0447	**13.0445**	13.0450	13.0453	13.0367	13.0416	13.0397	13.0395		
		0.5	13.0308	**13.0445**	13.0450	13.0463	13.0584	13.0360	13.0378	13.0380	13.0411		
		0.75	**13.0286**	13.0446	13.0454	13.0620	13.0816	**13.0339**	13.0381	13.0403	13.0432		
		1										**13.0391**	13.1884

Note: The smallest criterion value of each method and each quantile is in bold, indicating the best model for the corresponding method and quantile.

To check how the methods perform compared with a more parsimonious ℓ_1-type penalty, and how the results change with a different selection criterion, we estimate the Lasso model. The tuning parameter is chosen via both the BIC-type selection criterion and cross validation. Table 4 shows that all three methods perform better than Lasso for all three quantiles, except that adaptive group bridge with group bridge as the initial estimator performs worse than Lasso with BIC-type selection criterion for the 0.25 and 0.31 quantile. Results show that the Lasso with BIC-type selection criterion selects around 5 variables for all

three quantiles, less than half of the three methods. The selected variables belong to register AKM, BEF, IND, and LON. They are not entirely subsets of the selected registers of the three methods, but share similar information like the above-mentioned financial variables. The Lasso with cross validation selects 50 to 150 variables for all three quantiles, much more than the three methods (see also Table 5). The post-Lasso type inference shows that more than 60% of the selected variables are insignificant, showing evidence for the over selection pattern of cross validation. These results are unreported but available on request.

The choices of the tuning parameters play an important role in the selection process. Figure 3 shows the BIC-type criterion value for 100 values of ξ_N. The corresponding model is adaptive group bridge using competing risks quantile regression as the initial estimator with $(\gamma, \nu) = (0.75, 0.5)$ for the 0.25 quantile. There exists a global minimum at $\xi_N^* = 29.07$, and the corresponding BIC-type criterion value is 13.2431. In this way, we obtain one data point in Table 4. We run this process for all possible models. As shown in Table 4, the differences in criterion values are small for different combinations of (γ, ν) for each method and each quantile. The curve in Figure 3 is also rather flat for the range of ξ_N between 20 and 40, which suggests that the identification of the optimal ξ_N^* for the corresponding model is not strong either.

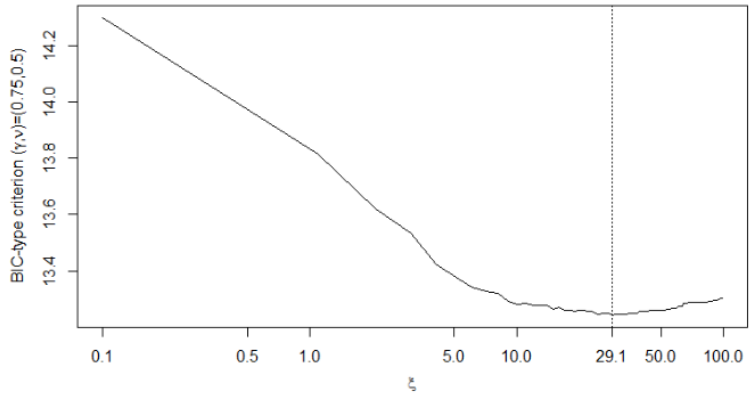

Fig.3 BIC-Type Criterion Value of AGB-CRQR for the 0.25 Quantile

Note: The corresponding model is AGB-CRQR with $(\gamma, \nu) = (0.75, 0.5)$. The dotted line indicates the smallest criterion value.

To check whether the results are rather insensitive to different combinations of (γ, ν), and different values of ξ_N, we add two robustness checks. Table 5 contains the overlap fraction of the selection results. The reference model is the optimal one according to the BIC-type criterion value for each method and each quantile as shown in Table 4. The fraction is calculated as the number of covariates that are both selected or both unselected between the reference model and the model considered divided by the total number of the covariates, which is 397. We see that the overlap fractions for the three methods and three quantiles are around 99%, which are very high. Table 6 contains the selection results with estimated coefficients at 20 values of ξ_N between 20 and 40 for the model in Figure 3. Only variables selected by at least one value of ξ_N are shown in the table. Results show that the selected variables are the subsets of the previous ones as ξ_N increases. The selected variable set changes slowly over the range of ξ_N, and the selected variables at the optimal ξ_N^* are selected at almost all values of ξ_N between 20 and 40. The signs of the estimates stay the same over the range of ξ_N, while the magnitudes decrease for most variables. According to Tables 5 and 6, the selection results are rather robust to different combinations of (γ, ν)

for all three methods and quantiles and different values of ξ_N in the rather flat area of the BIC-type criterion value curve, supporting that the optimal tuning parameters are rather reliable.

Table 5 Overlap Fraction for Various Models

			GB	AGB-GB				AGB-CRQR				LASSO	
				Tuning Parameter ν									
				BIC-Type Selection Criterion									CV
			0	0.5	1	1.5	2	0.5	1	1.5	2	0	0
Tuning Parameter γ	0.11 Quantile	0.25	0.990	0.982	0.980	0.980	0.992	0.990	0.980	0.985	0.982		
		0.5	0.992	0.982	0.987	0.992	0.982	0.990	0.987	0.997	0.982		
		0.75	**1.000**	**1.000**	0.990	0.992	0.987	0.977	0.997	**1.000**	0.990		
		1										**1.000**	0.724
	0.25 Quantile	0.25	0.992	0.985	0.990	0.985	0.987	0.995	0.992	0.987	0.987		
		0.5	**1.000**	0.982	**1.000**	0.995	0.990	1.000	0.992	0.987	0.987		
		0.75	1.000	0.990	0.997	0.985	0.987	**1.000**	0.987	0.985	0.985		
		1										**1.000**	0.686
	0.31 Quantile	0.25	0.990	1.000	**1.000**	1.000	1.000	0.995	0.987	0.990	0.990		
		0.5	1.000	**1.000**	1.000	1.000	0.990	1.000	0.992	0.987	0.990		
		0.75	**1.000**	1.000	1.000	0.985	0.997	**1.000**	0.992	0.987	0.985		
		1										**1.000**	0.874

Note: Bold indicates the reference model of each method and each quantile.

We refer once again to the post-Lasso type competing risks quantile regression model for an overview of the determinants of retirement transitions. We estimate a group bridge model with $(\gamma, \nu) = (0.5,1)$ at a grid of quantiles that are equally spaced from 0.01 to the upper bound with a step size of 0.01. We find that for $\tau > 0.31$ quantile, only the variable *Date of birth* is selected. Using *Date of birth* as the only variable, τ_U is 0.85. Therefore, if we use selected variables only at each corresponding quantile for the post-Lasso type estimation, we obtain estimates for quantiles between 0.01 and 0.85. The number of selected variables for each quantile is between 1 and 15, and the number of selected registers is between 1 and 3. We report estimated coefficients with 95% bootstrap confidence intervals in Figures S2.a and S2.b in Supplementary Material S.7. It is apparent that plots look very incomplete for some variables as the selection procedure jumps between (strongly correlated) variables when τ varies and most variables are unselected for $\tau > 0.3$. We therefore report the results for the union of the selected variables over all quantiles. The union includes 20 variables from 3 registers – labor market register AKM, population register BEF, and income register IND. τ_U is 0.58 when we use the union of the selected variables. This is below 0.85, resulted from the conditional nature of condition C4 in Peng and Fine (2009) (see Section 5).

In the following, we report estimated coefficients with 95% asymptotic confidence intervals, which are displayed in Figures 4a and 4b. In addition, we report estimated coefficients with 95% bootstrap confidence intervals in Figures S.1a and S.1b in Supplementary Material S.6 for comparison and as a robustness check. When comparing them it actually turns out that the two are very similar for quantiles lower than 0.50. For higher quantiles, however, the bootstrap confidence intervals explode, which could be due to violations of regularity conditions required for the validity of the bootstrap.

Figures 4a and 4b show the estimated coefficients of the 20 variables at a grid of quantiles that is equally spaced on [0.01, 0.58] with a step size of 0.01. The magnitudes, signs and significance levels change with τ for most variables. For the labor market register AKM, *Occupation: Operation/Transport*, *Occupation:*

Table 6 Estimation Results of AGB-CRQR for the 0.25 Quantile

	Within-Register Variable	Tuning Parameter ξ_N																			
		20.08	21.08	22.08	23.08	24.08	25.08	26.07	27.07	28.07	**29.07**	30.07	31.07	32.07	33.07	34.07	35.07	36.06	37.06	38.06	39.06
AKM	Occupation: Manual	-0.301	-0.270	-0.263	-0.256	-0.250	-0.230	-0.228	-0.221	-0.217	**-0.214**	-0.181	-0.175	-0.167	-0.168	-0.156	-0.138	-0.107	-0.109	-0.099	
	Occupation: Professional	0.350	0.326	0.336	0.330	0.339	0.341	0.344	0.340	0.344	**0.340**	0.362	0.348	0.346	0.347	0.351	0.351	0.297	0.278	0.283	
	Municipal employment	-0.218	-0.220	-0.218	-0.224	-0.228	-0.226	-0.228	-0.227	-0.205	**-0.199**	-0.180	-0.176	-0.175	-0.170	-0.152	-0.146	-0.125	-0.121	-0.121	
	Industry: Education	-0.048	-0.029	-0.028	-0.001																
	Industry: Healthcare	-0.040	-0.034	-0.031	-0.020	-0.020	-0.020	-0.018	-0.014	-0.035	**-0.036**	-0.042	-0.037	-0.039	-0.040	-0.039	-0.039	-0.034	-0.027	-0.016	
	Employed: Basic level	-0.308	-0.292	-0.289	-0.275	-0.272	-0.268	-0.267	-0.267	-0.257	**-0.258**	-0.242	-0.237	-0.236	-0.233	-0.216	-0.205	-0.191	-0.190	-0.181	
BEF	Date of birth	0.012	0.011	0.011	0.010	0.010	0.010	0.009	0.009	0.009	**0.009**	0.009	0.008	0.008	0.008	0.008	0.007	0.006	0.005	0.005	
	Divorced	0.091	0.053																		
	Family reference person	-0.103	-0.093	-0.086	-0.079	-0.075	-0.060	-0.051	-0.049	-0.035	**-0.040**	-0.025	-0.030	-0.029	-0.030	-0.032	-0.020				
	Male	0.012	0.015																		
IND	AM-income (mil.)	0.901	0.900	0.897	0.898	0.915	0.899	0.905	0.899	0.910	**0.900**	0.950	0.958	0.963	0.959	0.948	0.952	0.980	0.978	0.965	1.154
	ATP contributions (thous.)	0.009	0.004	0.009	0.002																
	Contributions to PEW (thous.)	-0.193	-0.196	-0.198	-0.198	-0.197	-0.202	-0.200	-0.201	-0.203	**-0.203**	-0.201	-0.201	-0.198	-0.198	-0.202	-0.202	-0.212	-0.214	-0.211	-0.209
	Taxable capital income (mil.)	-0.324	-0.158	-0.079	-0.140	-0.077															
	Total capital income (mil.)	0.457	0.542	0.609	0.649	0.590	0.533	0.538	0.524	0.403	**0.436**	0.213	0.153	0.168	0.167	0.156	0.170	0.186	0.190	0.215	0.662
	Pension income (mil.)	0.308	0.327	0.281	0.258	0.128	0.178	0.198	0.196	0.167											
	Total interest expense (mil.)	0.056	0.067	0.058	0.069	0.060	0.049	0.054	0.058	0.060											

Note: The model considered is AGB-CRQR with $(\gamma, v) = (0.75, 0.5)$. Bold indicates the optimal tuning parameter ξ_N according to the BIC-type selection criterion as shown in Figure 3.

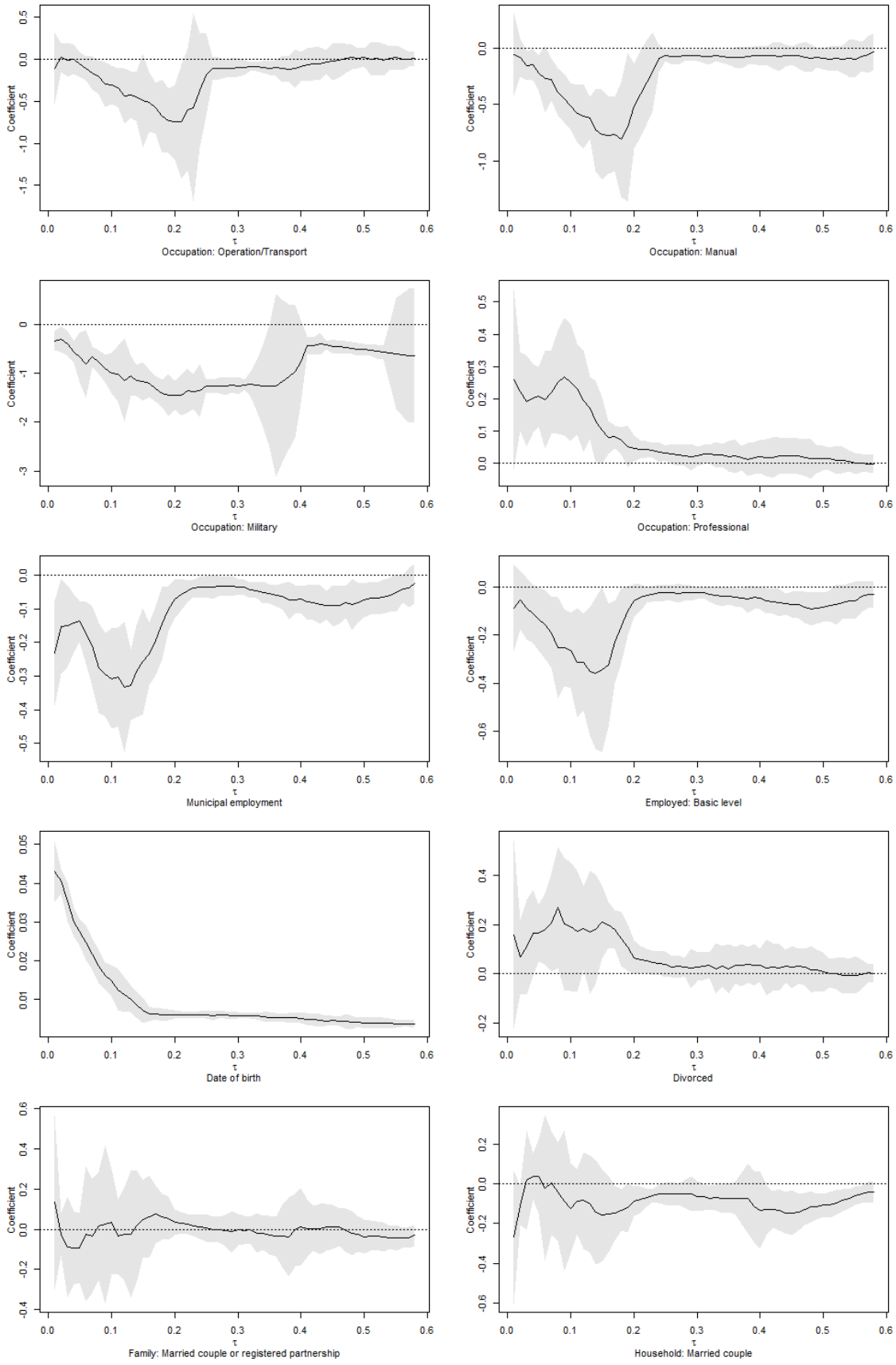

Fig.4a Post-Lasso Estimated Coefficients With 95% Asymptotic Confidence Interval for GB

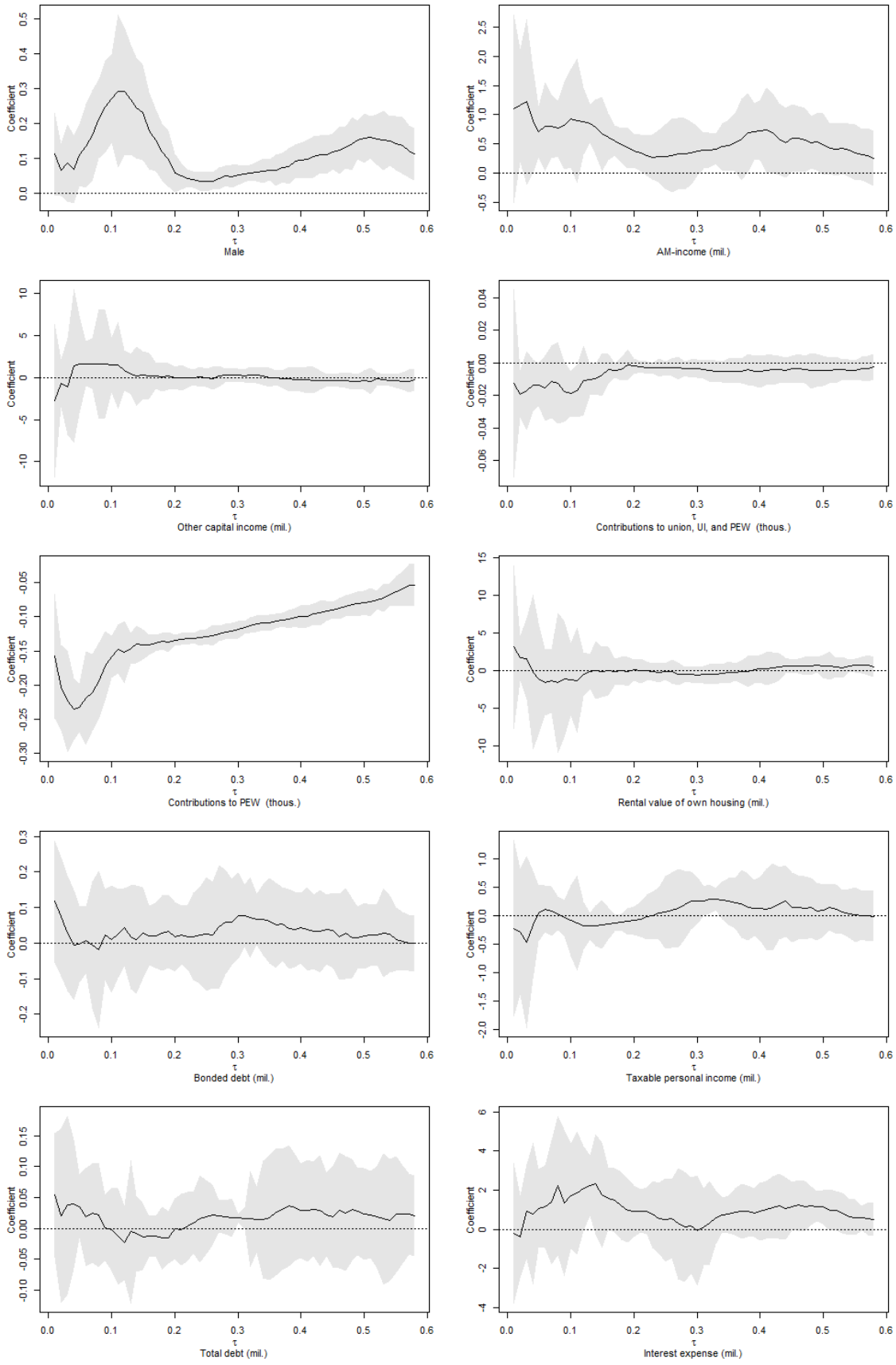

Fig.4b Post-Lasso Estimated Coefficients With 95% Asymptotic Confidence Interval for GB

Manual and *Occupation: Professional* are significant only for lower quantiles, while *Occupation: Military* and *Municipal employment* are significant for almost all quantiles. *Employed: Basic level* is significant for lower and the higher quantiles, not the middle quantiles. The signs do not change, and the magnitudes first increase and then decrease with τ for most variables. For the population register BEF, 5 variables are included in the union. Among them, *Family: Married couple* or *registered partnership* and *Household: Married couple* are highly correlated – more than 95% of the observations share the same values. They are selected by group bridge at different quantiles. So taking the union reduces the significance levels of these two variables to some extent. *Divorced* is significant for only lower quantiles, while *Date of birth* and *Male* are significant for almost all quantiles. The effects of *Date of birth* and *Divorced* are decreasing for most quantiles, suggesting that people with lower employment durations with exit to retirement are more sensitive to age requirement and marital status conditional on the covariates, and the longer the durations, the smaller the effects. On the contrary, *Male* has an increasing effect for people with both shorter and longer durations. For the income register IND, as mentioned earlier, multicollinearity is more prevalent. *AM-income* and *Taxable personal income* have a pairwise correlation around 0.93. The pairwise correlations among *Other capital income*, *Rental value of own housing*, *Bonded debt*, *Total debt*, and *Interest expense on mortgage debt* are from 0.4 to 0.95. *Contributions to union, unemp. insurance, and PEW* and *Contributions to PEW* have a pairwise correlation around 0.65. Comparing the estimation results with Figures S.2a and S.2b in Supplementary Material S.7, where only the selected variables are used at the corresponding quantiles, we find that taking the union considerably reduces the significance levels for almost all variables – most highly correlated variables are insignificant at almost all quantiles. *AM-income* is significant for the lower and middle quantiles, and *Contributions to PEW* is significant for all quantiles. The effect of *Contributions to PEW* is decreasing for the middle and higher quantiles, suggesting that contributions to early retirement scheme matter less for people who retire later ceteris paribus. Income related and debt related variables have positive effects on observed retirement transitions for most quantiles, same as Table 3.

Comparing Figures 4a and 4b with Figures S.2a and S.2b, we also find that group bridge does not select some variables that are significant if they are included in the post-Lasso type estimation at the corresponding quantiles. For example, the variable *Occupation: Manual* is significant at most quantiles between 0.08 and 0.36 in the post-Lasso type estimation, yet it is only selected at quantiles from 0.11 to 0.14, 0.23 and 0.24 by group bridge. The selection does not work well for quantiles between 0.16 and 0.22, because only less than or equal to 5 variables are selected at these quantiles, less than half of the numbers for the quantiles nearby. It could be due to the steepness or flatness of the cumulative incidence as discussed in Section 5. In all, we think that the selection results at each quantile indicate the important variables for the question in mind, yet it is safer to use the union of all selected variables for the post-Lasso type estimation, as the selection could be affected by many factors. The disadvantage is that the significance levels may be considerably reduced due to high multicollinearity.

Table 7 shows estimated trimmed mean effects and the constant test for these variables. The sign of a trimmed mean estimate shows the direction of the respective covariate effect on average over the quantiles. For example, the negative sign of *Occupation: Manual* suggests that individuals with manual occupation on average tend to have ceteris paribus shorter employment durations with exit to retirement. Most of the signs are the same as in Table 3. The trimmed mean effects for almost all variables in the labor market register AKM and population register BEF are highly significant, while more than half of the financial variables in the income register IND are not. The constant test suggests that most variables in AKM have

varying effects over quantiles, while the opposite is the case for BEF and IND. The results are in line with the graphical patterns of the coefficient estimates in Figures 4a and 4b – the estimates change either significantly or insignificantly across quantiles. For example, the test result and graph on *Occupation: Manual* show that having a manual occupation significantly shortens the employment durations with exit to retirement only at shorter durations. At longer durations, this effect becomes insignificant. Because the variables of the BEF and IND registers possess multicollinearity patterns, the resulting estimates show a lack of significance and it is difficult to trace out the covariate effects sharply. A failure to reject a trimmed-mean effect or a constant effect should be therefore interpreted with some caution in these cases. Overall, the results support the relevance of the quantile regression model in providing detailed information on the heterogeneous effects of covariates on the conditional quantiles of cumulative incidences.

Table 7 Trimmed Mean Effects and Constant Test Results

Register	Within-Register Variable	Trimmed Mean Effect		Constant Test
		Est.	P Value	P Value
AKM	Occupation: Operation/Transport	-0.186	0.023	0.003
	Occupation: Manual	-0.223	0.000	0.000
	Occupation: Military	-0.894	0.000	0.130
	Occupation: Professional	0.073	0.001	0.000
	Municipal employment	-0.112	0.000	0.000
	Employed: Basic level	-0.102	0.001	0.018
BEF	Date of birth	0.009	0.000	0.000
	Divorced	0.073	0.024	0.005
	Family: Married couple or registered partnership	-0.005	0.910	0.630
	Household: Married couple	-0.085	0.030	0.750
	Male	0.117	0.000	0.380
IND	AM-income (mil.)	0.562	0.002	0.370
	Other capital income (mil.)	0.109	0.800	0.370
	Contributions to union, unemp. insurance, and PEW (thous.)	-0.006	0.042	0.250
	Contributions to PEW (thous.)	-0.123	0.000	0.000
	Rental value of own housing (mil.)	0.044	0.940	0.640
	Bonded debt (mil.)	0.031	0.380	0.840
	Taxable personal income (mil.)	0.045	0.790	0.180
	Total debt (mil.)	0.016	0.460	0.690
	Interest expense on mortgage debt (mil.)	0.939	0.095	0.620

Figure 5 shows the kernel density estimates of the implied conditional quantiles of the cumulative incidence function for each quantile from 0.01 to 0.58 for all individuals in the sample. We use the estimates from the post-Lasso type estimation to compute the estimated conditional quantiles. Panel A shows the difference (in weeks) between the estimated conditional quantiles and the duration from the first week of 2009 until one reaches age of 65. Remember that 65 is the official retirement age, so a positive value indicates that an individual is estimated to retire late, i.e. to retire after the official retirement age, and a negative value indicates that an individual is estimated to retire early, i.e. to retire before the official retirement age. Panel B shows the estimated conditional quantiles. The estimated conditional quantiles cover the range of observed durations, where the longest durations in the sample are right-censored and are 418 weeks (compare Figure 1). The higher the quantiles, the more the curves are to the right, where we observe longer employment durations with exit to retirement. The estimated conditional quantiles for the lower

and middle quantiles capture the first two peaks in Figure 1. The estimated conditional quantiles for the higher quantiles not only capture the third peak, but also cover the right-censored durations in the sample, where individuals are still employed and not retired yet. Additionally, we provide Figure S.3 in Supplementary Material S.8, where only the selected variables at each quantile are used to compute the estimated conditional quantiles at the corresponding quantiles. In this case, the model is estimated for τ from 0.01 to 0.85. The results are quite different. We no longer obtain fitted values that correspond to late retirement, as the differences between the estimated conditional quantiles and the duration until one reaches age of 65 are centered at zero for quantiles from 0.47 to 0.85 (Panel A of Figure S.3). At the same time, the estimated conditional quantiles are concentrated at around 300 weeks for these quantiles (Panel B of Figure S.3). The strongly reduced variation in the fitted values is because the selected variable sets become smaller and smaller for the higher quantiles. Overall, this also points to the results for the union of selected variables being more convincing, because the estimated conditional quantiles cover the full range of the observed durations, and some unselected variables appear significant at the corresponding quantiles.

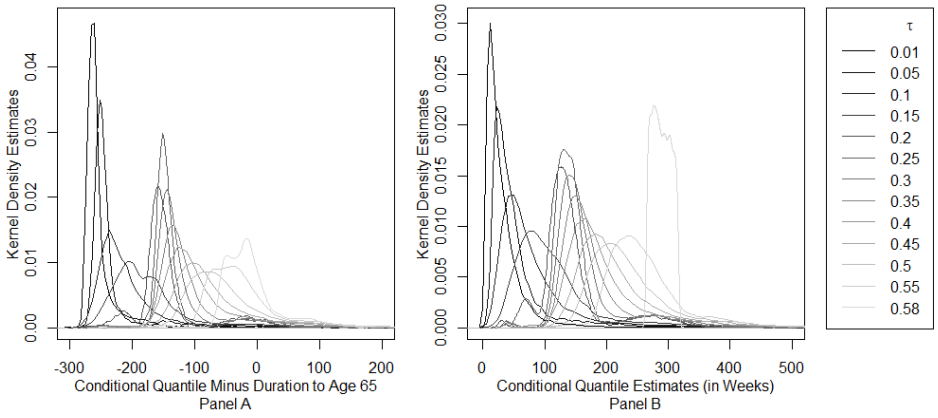

Fig.5 Distribution of Fitted Conditional Quantiles

7 Conclusions and Discussion

We present how machine learning, in particular, (adaptive) group bridge is helpful for bi-level variable selection in the competing risks quantile regression. To our knowledge, this is the first application of (adaptive) group bridge variable selection techniques in economics or social sciences and the first application of (adaptive) group bridge with competing risks quantile regression in high-dimensional linked administrative data. A union of 6 out of 16 registers and 32 out of 397 within-register variables are selected in Table 3 and 7. The selected variables contain demographic, socioeconomic, financial, and labor market information, have reasonable interpretation, and are significant in the unpenalized competing risks quantile regression. From the competing risks quantile regression model, we find that the magnitudes and significance levels of most estimated coefficients change strongly across quantiles, suggesting heterogeneous effects on transitions from employment into retirement for different durations and thus different retirement programs. All in all our results appear plausible and suggest that the (adaptive) group bridge can drastically reduce the dimensionality of the model, while maintaining a good model fit. We therefore suggest it should be included in the statistical toolbox of applied economic research. Our results should be also of interest to data providers of similar data. Linked administrative data are highly confidential and only access to

relevant information should be granted. Our results suggest that many highly sensitive variables are unselected and therefore actually do not contribute to the analysis. Therefore, data providers could use such tools in the initial stages of a project to restrict data access to relevant pieces of information.

Although, we do not analyze other exit routes out of employment, the quantile regression results provide differentiated evidence for the role of variables to change with retirement program. If we used a mean regression or a binary response model, the diversity of effects could not be revealed. In our analysis, most people enter into either early retirement pension or old age pension, which are clearly separated in terms of durations, and the quantile regression technique enables us to distinguish the heterogeneous effects of variables on different cumulative incidence quantiles.

However, we do face some challenges. The potential multicollinearity and misclassification in our data and selection issues in data preparation possibly affect our results. Another problem associated with multicollinearity is that we cannot trace out the role of a register if relevant within-register variables are dropped due to their high correlation with variables from other registers. (Adaptive) group bridge allows the same variable to appear in different groups, but in our case, we have different but highly correlated variables. So, additional work is required to handle this problem. To generalize the validity of the bootstrap inference, it would be of interest to carry over the wild bootstrap procedure of Wang et al. (2018) to the penalized competing risks quantile regression model.

List of References

Ahn, K. W., Banerjee, A., Sahr, N., & Kim, S. (2018). Group and within-group variable selection for competing risks data. Lifetime data analysis, 24(3), 407-424.

Ahn, K.W., & Kim, S. (2018). Variable selection with group structure in competing risks quantile regression. Statistics in Medicine, 37(9), 1577-1586.

Amilon, A., & Nielsen, T. H. (2010). How does the option to defer pension payments affect the labour supply of older workers in Denmark? Working and Ageing: Emerging Theories and Empirical Perspectives, 190-209. European Centre for the Development of Vocational Training.

An, M., Christensen, B., & Gupta, N. (2004). Multivariate mixed proportional hazard modelling of the joint retirement of married couples. Journal of Applied Econometrics, 19(6), 687-704.

Barslund, M. (2015). Extending working lives: The case of Denmark. CEPS Working Document, No. 404. Available at SSRN: https://ssrn.com/abstract=2577739.

Belloni, A., & Chernozhukov, V. (2011). ℓ1-penalized quantile regression in high-dimensional sparse models. The Annals of Statistics, 39(1), 82-130.

Belloni, A., & Chernozhukov, V. (2013). Least squares after model selection in high-dimensional sparse models. Bernoulli, 19(2), 521-547.

Bingley, & Lanot. (2007). Public pension programmes and the retirement of married couples in Denmark. Journal of Public Economics, 91(10), 1878-1901.

Bingley, P., Gupta, N. D., & Pedersen, P. J. (2004). The Impact of Incentives on Retirement in Denmark. In J. Gruber, & D. Wise (Eds.), Social Security Programs and Retirement Around the World: Microestimation (pp. 153-234). Chicago: University of Chicago Press. A National Bureau of Economic Research conference report.

Bingley, P., Gupta N.D. and Pedersen, P.J. (2012). Disability Programs, Health, and Retirement in Denmark since 1960. NBER Chapters, in: Social Security Programs and Retirement around the World: Historical Trends in Mortality and Health, Employment, and Disability Insurance Participation and Reforms, 217--249 National Bureau of Economic Research, Inc.

Bingley, P., Gupta, N. D., Jørgensen, M., & Pedersen, P. J. (2016). Health, Disability Insurance, and Retirement in Denmark. In D. A. Wise (Ed.), Social Security Programs and Retirement around the World: Disability Insurance Programs and Retirement Chicago: University of Chicago Press.

Chatterjee, A., & Lahiri, S. N. (2011). Bootstrapping Lasso estimators. Journal of the American Statistical Association, 106(494), 608-625.

Chatterjee, A., & Lahiri, S. N. (2013). Rates of convergence of the adaptive LASSO estimators to the oracle distribution and higher order refinements by the bootstrap. The Annals of Statistics, 41(3), 1232-1259.

Chen, J. E., & Tien, J. J. (2019). Debiased/Double Machine Learning for Instrumental Variable Quantile Regressions. arXiv.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., & Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. Econometrics Journal, 21(1), C1-C68.

Christensen, B., & Kallestrup-Lamb, M. (2012). The Impact of Health Changes on Labor Supply: Evidence from merged Data on Individual Objective Medical Diagnosis Codes and Early Retirement Behavior. Health Economics, 21(Supp1), 56-100.

Cox, D.R. (1962), Renewal Theory, London.

Datta Gupta, N., & Larsen, M. (2007). Health shocks and retirement: The role of welfare state institutions. European Journal of Ageing, 4(3), 183-190.

Datta Gupta, N., & Larsen, M. (2010). The impact of health on individual retirement plans: Self-reported versus diagnostic measures. Health Economics, 19(7), 792-813.

Dlugosz S., Peng, L., & Li, R. (2019). cmprskQR: Analysis of Competing Risks Using Quantile Regressions. R package version 0.9.2. https://CRAN.R-project.org/package=cmprskQR

Dlugosz, S., Lo, S., & Wilke, R. (2017). Competing risks quantile regression at work: In-depth exploration of the role of public child support for the duration of maternity leave. Journal of Applied Statistics, 44(1), 109-122.

Duval, R. (2003). The Retirement Effects of Old-age Pension and Early Retirement Schemes in OECD Countries. OECD Economics DepartmentWorking Paper 370, OECD.

Emura, T., Shih, J.H., Ha, I.D. and Wilke, R.A. (2019) Comparison of the marginal hazard model and the subdistribution hazard model for competing risks under an assumed copula. Statistical Methods in Medical Research, 0(0), 1-21.

Fermanian, J. (2003). Nonparametric estimation of competing risks models with covariates. Journal of Multivariate Analysis, 85(1), 156-191.

Filges, T., Larsen, M. and Pedersen, P. J. (2012). Retirement: Does Individual Unemployment Matter? Evidence from Danish Panel Data 1980–2009. IZA Discussion Paper, No. 6538, Institute for the Study of Labor (IZA).

Fine, J. P., & Gray, R. J. (1999). A proportional hazards model for the subdistribution of a competing risk. Journal of the American statistical association, 94(446), 496-509.

Fitzenberger, B., & Wilke, R. A. (2010a). New insights into unemployment duration and post unemployment earnings in Germany. Oxford Bulletin of Economics and Statistics, 72(6), 794-826.

Fitzenberger, B., & Wilke, R. (2010b). Unemployment Durations in West Germany Before and After the Reform of the Unemployment Compensation System during the 1980s. German Economic Review, 11(3), 336-366.

Fu, W. J. (1998). Penalized regressions: the bridge versus the Lasso. Journal of computational and graphical statistics, 7(3), 397-416.

Gruber, J., & Wise, D. (1998). Social Security and Retirement: An International Comparison. The American Economic Review, 88(2), 158-163.

Hastie, T., Tibshirani, R., & Wainwright, M. (2015). Statistical learning with sparsity: the Lasso and generalizations. Chapman and Hall/CRC.

Heckman, J. and Honoré, B. (1989) The identifiability of the competing risks model. Biometrika, 76(2), 325-330.

Hoerl, Arthur E. (1962). Application of Ridge Analysis to Regression Problems. Chemical Engineering Progress, 58(3), 54-59.

Huang, J., Breheny, P., & Ma, S. (2012). A Selective Review of Group Selection in High-Dimensional Models. Statistical Science, 27(4), 481-499.

Huang, J., Liu, L., Liu, Y., & Zhao, X. (2014). Group selection in the Cox model with a diverging number of covariates. Statistica sinica, 1787-1810.

Huang, J., Ma, S., Xie, H., & Zhang, C. (2009). A group bridge approach for variable selection. Biometrika, 96(2), 339-355.

Kallestrup-Lamb, M., Kock, A., & Kristensen, J. (2016). Lassoing the Determinants of Retirement. Econometric Reviews, 35(8-10), 1-40.

Koenker, R., & Bassett, G. (1978). Regression Quantiles. Econometrica, 46(1), 33-50.

Koenker, R., & Bilias, Y. (2001). Quantile regression for duration data: A reappraisal of the Pennsylvania Reemployment Bonus Experiments. Empirical Economics, 26(1), 199-220.

Koenker, R., Chernozhukov, V., He, X., & Peng, L. (Eds.). (2017). Handbook of quantile regression. Chapman and Hall/CRC press.

Koenker, R., & Geling, O. (2001). Reappraising medfly longevity: a quantile regression survival analysis. Journal of the American Statistical Association, 96(454), 458-468.

Koenker, R. (2017). quantreg: Quantile Regression. R package version 5.33. https://CRAN.R-project.org/package=quantreg

Kyyrä, T., & Wilke, R. (2007). Reduction in the Long-term Unemployment of the Elderly: A Success Story from Finland. Journal of the European Economic Association, 5(1), 154-182.

Larsen, M., & J Pedersen, P. (2005). Pathways to Early Retirement in Denmark, 1984-2000. IZA Discussion Paper, No. 1575, Institute for the Study of Labor (IZA).

Larsen, M., & Pedersen, P. (2013). To work, to retire – or both? Labor market activity after 60. IZA Journal of European Labor Studies, 2(1), 1-20.

Lee, J. D., Sun, D. L., Sun, Y., & Taylor, J. E. (2016). Exact post-selection inference, with application to the Lasso. The Annals of Statistics, 44(3), 907-927.

Lindeboom, M. (1998). Microeconometric Analysis of the Retirement Decision: The Netherlands. OECD Economics Department Working Papers, No. 207, OECD Publishing.

Lockhart, R., Taylor, J., Tibshirani, R. J., & Tibshirani, R. (2014). A significance test for the Lasso. Annals of statistics, 42(2), 413.

Machado, J. A. F., & Silva, J. S. (2005). Quantiles for counts. Journal of the American Statistical Association, 100(472), 1226-1237.

Meinshausen, N., Meier, L., & Bühlmann, P. (2009). P-values for high-dimensional regression. Journal of the American Statistical Association, 104(488), 1671-1681.

Miniaci, R. and Stancanelli, E. (1998). Microeconometric Analysis of the Retirement Decision: United Kingdom. OECD Economics Department Working Papers, No. 206, OECD Publishing.

Møller Danø, Ejrnæs, & Husted. (2005). Do single women value early retirement more than single men? Labour Economics, 12(1), 47-71.

OECD (2012a). Thematic Follow-up Review of Policies to Improve Labour Market Prospects for Older Workers: Denmark. OECD Publishing.

Peng, L., & Fine, J. (2009). Competing Risks Quantile Regression. Journal of the American Statistical Association, 104(488), 1440-1453.

Peterson, A. V. (1976). Bounds for a joint distribution function with fixed sub-distribution functions: Application to competing risks. Proceedings of the National Academy of Sciences, 73(1), 11-13.

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

Statistics Denmark. (2016). Documentation of statistics for Register-Based Labour Force Statistics. [online] Available at: https://www.dst.dk/en/Statistik/dokumentation/documentationofstatistics/register-based-labour-force-statistics.

Taylor, J., & Tibshirani, R. J. (2015). Statistical learning and selective inference. Proceedings of the National Academy of Sciences, 112(25), 7629-7634.

Tibshirani, R. (1996). Regression Shrinkage and Selection Via the Lasso. Journal of the Royal Statistical Society: Series B (Methodological), 58(1), 267-288.

Van de Geer, S., Bühlmann, P., Ritov, Y. A., & Dezeure, R. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. The Annals of Statistics, 42(3), 1166-1202.

Wang, L., Van Keilegom, I., & Maidman, A. (2018). Wild residual bootstrap inference for penalized quantile regression with heteroscedastic errors. Biometrika, 105(4), 859-872.

Yuan, M., & Lin, Y. (2006). Model selection and estimation in regression with grouped variables. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 68(1), 49-67.

Zhang, C. H., & Zhang, S. S. (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 76(1), 217-242.

Zou, H. (2006). The Adaptive Lasso and Its Oracle Properties. Journal of the American Statistical Association, 101(476), 1418-1429.

Appendix 1

Complete algorithm for the (adaptive) group bridge:

1. Choose a certain quantile.
2. Set up a 4-dimensional grid $\mathcal{G} = \xi_N \times \gamma \times \nu \times \tilde{\beta} = \xi_N \times \{0.25, 0.5, 0.75\} \times \{0, 0.5, 1, 1.5, 2\} \times \tilde{\beta}$ for the tuning parameters. Following Friedman et al. (2010), we choose 100 uniformly spaced values for ξ_N. The upper bound is the smallest value where none of the variables is selected and the lower bound is the upper bound divided by 1000. For the initial estimator $\tilde{\beta}$, we use the group bridge estimator or the unpenalised competing risks quantile regression estimator to compute the individual weights for the adaptive group bridge.
3. Choose one grid point of the tuning parameters $(\gamma, \nu, \tilde{\beta})$. For each value of the tuning parameter ξ_N, repeat the following steps for $t = 1, \ldots$ until practical convergence indicated by $\left\|\hat{\beta}^t(\tau) - \hat{\beta}^{t-1}(\tau)\right\|_1 < 0.001$, and save the estimated coefficients $\hat{\beta}(\tau)$ after practical convergence:

   a) Compute $\theta_j^{(t)} = A_j^{1-\gamma}\left(\frac{1-\gamma}{\gamma}\right)^{\gamma}\left(\sum_{k=1}^{A_j}\left(\left|\beta_{jk}^{(t-1)}(\tau)\right| \Big/ \left|\tilde{\beta}_{jk}(\tau)\right|^{\nu}\right)\right)^{\gamma}$ for all groups $j = 1, \ldots, J$, where for the first iteration $\beta_{jk}^{(0)}(\tau) = \tilde{\beta}_{jk}(\tau)$.

   b) Solve the minimization problem of (adaptive) group bridge

$$\hat{\beta}^t(\tau) = \underset{b(\tau)}{argmin}\ U_N(b(\tau), \tau) + \xi_N \sum_{j=1}^{J}\left(\left(\frac{\theta_j^{(t)}}{A_j}\right)^{1-\frac{1}{\gamma}} \sum_{k=1}^{A_j}\left(\left|b_{jk}(\tau)\right| \Big/ \left|\tilde{\beta}_{jk}(\tau)\right|^{\nu}\right)\right),$$

$$= \underset{b(\tau)}{argmin}\ U_N(b(\tau), \tau) + \xi_N \sum_{j=1}^{J}\sum_{k=1}^{A_j} w_{jk}^{(t)}\left|b_{jk}(\tau)\right|$$

   where $w_{jk}^{(t)} = \left(\frac{\theta_j^{(t)}}{A_j}\right)^{1-\frac{1}{\gamma}} \times \frac{1}{\left|\tilde{\beta}_{jk}(\tau)\right|^{\nu}}$.

4. Now we have estimates $\hat{\beta}(\tau)$ for 100 values of the tuning parameter ξ_N. Then we compute the BIC-type criterion proposed by Ahn and Kim (2018),

$$\frac{2}{N} U_N(\hat{\beta}(\tau), \tau) + p_N\, ln(K)\, \frac{ln(N)}{2N}.$$

   Choose the optimal ξ_N that leads to the smallest criterion value and save the criterion value.
5. Repeat (3)-(4) for all grids points of $(\gamma, \nu, \tilde{\beta})$. The tuning parameters $(\xi_N^*, \gamma^*, \nu^*, \tilde{\beta}^*)$ that leads to the smallest criterion value gives the optimal estimates $\hat{\beta}(\tau)$.