I Z A

Penalized Quantile Regression with Semiparametric Correlated Effects: Applications with Heterogeneous Preferences

Matthew Harding
Carlos Lamarche

Penalized Quantile Regression with Semiparametric Correlated Effects: Applications with Heterogeneous Preferences

Matthew Harding
Stanford University

Carlos Lamarche
*University of Kentucky
and IZA*

ABSTRACT

Penalized Quantile Regression with Semiparametric Correlated Effects: Applications with Heterogeneous Preferences[*]

This paper proposes new ℓ_1-penalized quantile regression estimators for panel data, which explicitly allows for individual heterogeneity associated with covariates. We conduct Monte Carlo simulations to assess the small sample performance of the new estimators and provide comparisons of new and existing penalized estimators in terms of quadratic loss. We apply the techniques to two empirical studies. First, the new method is applied to the estimation of labor supply elasticities and we find evidence that positive substitution effects dominate negative wealth effects at the middle of the conditional distribution of hours. The overall effect tends to be larger at the lower tail, which suggests that changes in taxes have different effects across the response distribution. Second, we estimate consumer preferences for nutrients from a demand model using a large scanner dataset of household food purchases. We show that preferences for nutrients vary across the conditional distribution of expenditure and across genders, and emphasize the importance of fully capturing consumer heterogeneity in demand modeling. Both applications highlight the importance of estimating individual heterogeneity when designing economic policy.

JEL Classification: C21, C23, J22

Keywords: shrinkage, panel data, quantile regression, labor supply, scanner data

Corresponding author:

Matthew Harding
Department of Economics
Stanford University
579 Serra Mall
Stanford, CA 94305
USA
E-mail: mch@stanford.edu

---

1. Introduction

The recent availability of Big Data opens up the possibility of devising targeted economic policies that increase welfare by accounting for the broad individual heterogeneity in both characteristics and outcomes. At the same time, the large datasets make it possible to provide increased flexibility in the specification of econometric models. This paper provides a simple new approach to the estimation of models with heterogeneous marginal effects in panel data with time-invariant variables by allowing for a flexible specification of correlated individual effects in a quantile regression setting.

There is a growing theoretical and empirical interest on the estimation of a quantile panel data model, specially after Koenker (2004). For recent developments, see Abrevaya and Dahl (2008), Graham, Hahn, and Powell (2009), Harding and Lamarche (2009, 2013), Lamarche (2010), Galvao (2011), Canay (2012), Rosen (2012), Galvao, Lamarche and Lima (2013), and Chernozhukov, Fernández-Val, Hahn and Newey (2013). Koenker (2004) proposes to jointly estimate a vector of covariate effects and a vector of individual effects considering a class of penalized quantile regression estimators. The method uses an ℓ_1 penalty term to control the bias and variance of the estimates of the covariate effects. Lamarche (2010) obtains the minimum variance estimator in the class of ℓ_1-penalized estimators under stochastic independence between individual effects and covariates. While existing fixed effects approaches might suffer from the incidental parameters problem, recent important developments in nonseparable models estimate the effect of independent variables on quantiles of the response variable, conditional on time-varying variables. Chernozhukov, Fernández-Val, Hahn and Newey (2013) offer identification and estimation results of quantile effects in nonseparable models.

Consider the example of a labor supply model. Flexibility in specification and unobserved heterogeneity play a fundamental role in the estimation of labor supply models (Burtless and Hausman (1978), Hausman (1985), Blundell and Meghir (1986), Ziliak and Kniesner (1999), Blundell and MaCurdy (1999)). However, most of the research on the subject used maximum likelihood and instrumental variable methods. These approaches rely on strong distributional assumptions and/or are limited to a conditional mean view of the relationship among variables. Recent important attempts to provide a more flexible analysis includes nonparametric approaches, but unobserved heterogeneity is assumed to be drawn from known (Gaussian) distributions (see, e.g., van Soest, Das, and Gong (2002)) or is not accounted for (see, e.g., Kumar (2012)). Quantile regression for panel data offers a flexible alternative approach to conditional mean analysis that is efficient under non-Gaussian conditions.

However, neither the estimator of Koenker (2004) nor the penalized quantile regression estimator of Lamarche (2010) is well suited for estimation of labor supply models. In empirical labor supply specifications, wages are suspected to be endogenous because unobserved time-invariant preferences for work may be correlated with latent factors affecting productivity and hence wages (Blundell, MaCurdy and Meghir 2007).

The penalized quantile regression estimator can be extended to models with endogenous individual effects. In this paper, we propose penalized estimators that can be easily applied to a class of semiparametric models (Cai and Xiao (2012)) and parametric models (Abrevaya and Dahl (2008)). As in Koenker (2004), the individual effects represent location shift effects on the conditional quantiles of the response, and therefore, we avoid issues associated with estimating a quantile regression model with additive error terms (Koenker and Hallock (2000), Rosen (2012)). The estimation of these additional parameters increases the variability of the estimates of the covariate effects, but shrinkage can be used to control the additional variability. We use an ℓ_1 penalty term (Tibshirani (1996), Donoho et al. (1998)) to shrink a vector of individual effects and a tuning parameter λ to control the degree of this shrinkage. We present necessary conditions for our method to reduce the variability of the estimate of the slope parameter without sacrificing bias. The approach allows estimation of time-invariant taste shifters and wealth effects in an empirical labor supply model, and it is not more difficult to implement than other quantile regression panel data methods.

The penalized estimator proposed in this paper can be seen as a balanced compromise between misspecification issues arising from the omission of individual heterogeneity and the incidental parameters problem arising from leaving individual heterogeneity unrestricted in a nonlinear panel model. As first pointed out by Neyman and Scott (1948) and recently elaborated by Kato, Galvao and Montes-Rojas (2012), the estimation of individual effects in a nonlinear panel data model leads to inconsistent estimates of the slope parameters. Kato, Galvao and Montes-Rojas (2012) employs large T and N asymptotics offering restrictions on the growth of T which are unusual in micro-econometric panels but serve as important warning devices to practitioners. Under less general conditions, Graham, Hahn, and Powell (2009) show that there is no incidental parameter problem in a non-differentiable panel data model and Koenker (2004) and Galvao, Lamarche, Lima (2013) show empirical evidence that the bias of the fixed effects estimator is small for moderate T. On the other hand, as in the classical linear panel models described in detail in Hsiao (2003) and Baltagi (2008), ignoring unobserved heterogeneity generally leads to inconsistent estimates of the slope parameters. We show that the penalized estimator reduces the noise in the estimation of the individual effects while controlling for individual heterogeneity.

This paper seeks to contribute the literature by comparing the new and existing ℓ_1-penalized quantile regression estimators in terms of quadratic loss. We first show that Koenker's (2004) estimator is the efficient estimator in the class of panel data quantile regression estimators. However, the proposed approach has smaller asymptotic mean squared error than the penalized estimator if the correlation between independent variables and latent individual factors is not negligible. We also show that by choosing λ carefully, we can make the asymptotic mean squared error of the estimator smaller than the asymptotic mean squared error of a quantile regression estimator for the correlated effects model. This indicates that shrinking individual effects potentially uncorrelated with independent variables is worthwhile. We provide conditions under which the strictness of the penalization can be determined by minimizing mean squared error.

The next section presents the models and estimators. Section 3 derives the asymptotic mean squared error of a proposed estimator and Section 4 offers Monte-Carlo evidence. In Section 5, we first show that an empirical quantile regression model of labor supply can be simply motivated by allowing for heterogeneous preferences in the classical life-cycle framework of consumption and labor supply. We then demonstrate how the penalized estimator can be obtained and used to estimate quantile specific labor supply elasticities. We also provide an additional empirical example which addresses the problem of estimating preference heterogeneity in consumer demand models using scanner datasets, where differences between the preference distributions over product attributes vary by socio-demographics. Section 6 provides conclusions.

2. Models and Estimators

Let the data be observations $\{(y_{it}, \boldsymbol{x}'_{it}) : i = 1, ...N, t = 1, ...T\}$ from a random coefficient version of a quantile regression panel data model:

$$(2.1) \qquad y_{it} \;=\; \boldsymbol{x}'_{it}\boldsymbol{\beta}(u_{it}) + \alpha_i(u_{it})$$

$$(2.2) \qquad \tau \;\mapsto\; \boldsymbol{x}'_{it}\boldsymbol{\beta}(\tau) + \alpha_i(\tau)$$

where y_{it} is the dependent variable, $\boldsymbol{x}_{it} = (1, x_{it,2}, ..., x_{it,p})'$ is the vector of independent variables, the α_i's are unobservable time-invariant effects, $u_{it}|\boldsymbol{x}_{it}, \alpha_i \sim \mathcal{U}(0,1)$ denotes a uniform distribution, and τ is the τ-th quantile of the conditional distribution of the response variable. The right hand side of (2.2) is the conditional quantile function, $Q_{Y_{it}}(\tau|\boldsymbol{x}_{it}, \alpha_i) = \inf\{y : \Pr(y_{it} < y|\boldsymbol{x}_{it}, \alpha_i) \geq \tau\}$ for all τ in $(0,1)$. The parameter $\boldsymbol{\beta}(\tau)$'s models the covariate effect providing an opportunity for investigating how time-variant observable factors

influence the location, scale and shape of the conditional distribution of the response. For simplicity, the model does not include time-invariant explanatory variables, which can be easily incorporated as shown in Section 5. It is also assumed that the panel is balanced, with observations $(y_{it}, \boldsymbol{x}'_{it})' \in \mathbb{R} \times \mathbb{R}^p$ for each of the N subjects over $t = 1, \ldots, T$.

The model takes a semiparametric form because no parametric assumption is made on the relationship between the vector of covariates $\boldsymbol{x}_{it}$ and α_i and the functional form of the conditional distribution of the response variable is left unspecified. The unobserved variable α_i could be arbitrarily related to observable variables and unobservable variables:

$$(2.3) \qquad \alpha_i(\boldsymbol{x}_i, u_{i1}, a_i) = g(\boldsymbol{x}_i, u_{i1}) + a_i,$$

where $g(\cdot)$ is an unknown function with a certain degree of smoothness, the independent variable $\boldsymbol{x}_i = (\boldsymbol{x}'_{i1}, \boldsymbol{x}'_{i2}, ..., \boldsymbol{x}'_{iT})'$ and the individual effect a_i is, by definition, uncorrelated with the independent variables. We allow the variables α_i and $\boldsymbol{x}_{it}$ to be stochastically dependent by considering the individual effect to be drawn from a conditional distribution function with location $g(\tau, \boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{iT})$.

It is important to note that (2.3) imposes a time-homogeneity condition similar to Assumption 2 in Chernozhukov, Fernández-Val, Hahn and Newey (2013). The implication of this is that the regressors are "strictly exogenous" with respect to a_i. At the same time, it requires that the conditional distribution of $u_{it}|\boldsymbol{x}_i, a_i$ does not depend on t (e.g., the distribution of $a_i, u_{it}|\boldsymbol{x}_i$ is identical to the distribution of $a_i, u_{i1}|\boldsymbol{x}_i$).

Under a monotonicy condition assumed in equation (2.2), the model (2.1)-(2.3) represents a more general version of several specifications recently proposed in the growing literature on quantile panel data models. Consider for instance the following variations of interest in the theoretical and empirical literature.

Example 1. If $\alpha_i(\boldsymbol{x}_i, u_{i1}, a_i) = g(u_{i1})$ for all i and $T = 1$, then model (2.1) and (2.3) becomes a semiparametric quantile regression model, $y_{i1} = \boldsymbol{x}'_{i1}\boldsymbol{\beta}(u_{i1}) + g(u_{i1})$, similar to the cross-sectional models investigated by He and Shi (1996) and Cai and Xiao (2012).

Example 2. Semiparametric models for longitudinal data are investigated in Wei and He (2006) and Wei, Pere, Koenker and He (2006). Under the assumption that repeated measurements are regularly observed over time, a version of their model arises by replacing $x_{it,2}$ by y_{it-1} in equation (2.1) and $\boldsymbol{x}_i$ by t_i in equation (2.3). The conditional quantile function is equal to:

$$Q_{Y_{it}}(\tau|t_i, y_{it-1}, \boldsymbol{x}_{it}) = g(\tau, t_i) + \beta_2(\tau)y_{it-1} + \boldsymbol{x}'_{it}\boldsymbol{\beta}_{-2}(\tau) + a_i,$$

where $\boldsymbol{\beta}_{-2}(\tau) = (\beta_1(\tau), \beta_3(\tau), ..., \beta_p(\tau))$ is a $p - 1$-dimensional vector.

Example 3. If $\alpha_i(\boldsymbol{x}_i, u_{i1}, a_i)$ is a known parametric function, the model can be seen within the classical framework proposed by Chamberlain (1982) leading to a representation of endogenous individual effects $\alpha_i(\tau, \boldsymbol{x}_i, a_i) = \boldsymbol{x}_i'\boldsymbol{\gamma}(\tau) + a_i$. Abrevaya and Dahl (2008) study estimation of a quantile regression model under the assumption that equation (2.3) is equal to $\boldsymbol{x}_i'\boldsymbol{\gamma}(\tau)$, and Koenker (2004), Lamarche (2010) and Canay (2011) study estimation of the model under the assumption $\alpha_i(\tau) = \alpha_i$ for all i.

2.1. Estimation procedures

Our estimation approaches for model (2.1) and (2.3) serve as an intermediate class of procedures with good robustness of possible deviations from the classical correlated random effects model and relatively more precise estimation of the parametric part of the quantile regression model.

The estimation procedure for the model with flexible correlated effects proceeds in two steps; see Cai and Xiao (2012), He and Shi (1996) and Tang, Wang, He and Zhu (2012) for a related discussion. First, we express $g(\tau, \boldsymbol{x}_i)$ as a linear expansion of B-splines. Although other nonparametric regression techniques can be used in a first stage, the linear formulation of the B-splines yields a family of quantile functions that can be easily accommodated to a quantile regression for panel data problem. We express,

$$(2.4) \qquad g(\boldsymbol{x}_i)'\boldsymbol{\gamma}(\tau) \approx \boldsymbol{b}(\boldsymbol{x}_{i1})'\boldsymbol{\gamma}_1(\tau) + \boldsymbol{b}(\boldsymbol{x}_{i2})'\boldsymbol{\gamma}_2(\tau) + ... + \boldsymbol{b}(\boldsymbol{x}_{iT})'\boldsymbol{\gamma}_T(\tau),$$

where $\boldsymbol{b}(\boldsymbol{x}_{ij}) = (b_1(\boldsymbol{x}_{ij}), \ldots, b_{k_n+h+1}(\boldsymbol{x}_{ij}))'$ is a B-spline basis function, k_n is the number of knots, h is the B-spline basis, and $\boldsymbol{\gamma}$ is the spline coefficient vector. We employ cubic B-spline basis functions with $k \propto (NT)^{1/5}$ with knots selected as the empirical quartiles of $\boldsymbol{x}_{ij}$. The model becomes a linear quantile regression model in all coefficients and can be estimated using the following estimator,

$$(2.5) \qquad \min_{\beta,\gamma,a \in \mathcal{B} \times \mathcal{G} \times \mathcal{A}} \sum_{j=1}^{J} \sum_{t=1}^{T} \sum_{i=1}^{N} \omega_j \rho_{\tau_j}(y_{it} - \boldsymbol{x}_{it}'\boldsymbol{\beta}(\tau_j) - \hat{g}(\boldsymbol{x}_i)'\boldsymbol{\gamma}(\tau_j) - a_i) + \lambda Pen(\boldsymbol{a}),$$

where $\rho_{\tau_j}(u) = u(\tau_j - I(u \leq 0))$ is the quantile loss function, ω_j is a relative weight given to the j-th quantile, J is the number of quantiles $\{\tau_1, \tau_2, ..., \tau_J\}$ to be estimated, and λ is the *Tikhonov* regularization parameter or tuning parameter. The function $Pen(\boldsymbol{a})$ is a penalty term that could be defined as $\|\boldsymbol{a} - \boldsymbol{a}^*\|_1$, where $\boldsymbol{a}^*$ may be close to the unknown location of the distribution. In equation (2.3), a_i has zero mean by definition, so we made use of this

information defining the penalty term as,

$$Pen(\boldsymbol{a}) = \|\boldsymbol{a}\|_1,$$

where $\|\boldsymbol{a}\|_1$ is the standard ℓ_1-norm defined as $\|\boldsymbol{a}\|_1 \equiv \sum_i |a_i|$.

The estimation of the individual effects increases the variability of the estimator of the slope parameter, but this penalty term that shrinks the fixed effects estimator of the a_i's toward zero helps to reduce the inflation effect without sacrificing bias. When the a's are exchangeable and drawn from a conditional distribution function with location zero, shrinkage that forces some individual specific effect estimates $\hat{a}$'s to be zero does not impose bias and affects the performance of the estimator of the parameter of interest $\boldsymbol{\beta}(\boldsymbol{\tau})$. There is an enormous amount of work in statistics and lately in econometrics dealing with shrinkage in a wide spectrum of problems (see, e.g., Koenker (2004), Horowitz and Lee (2007), Carrasco, Florens and Renault (2007), Chen (2007), Belloni and Chernozhukov (2011); see also Bickel and Li (2006) for a survey in statistics).

Although flexibility in specification is naturally important, Euler equations obtained from economic models are associated with practical choices. We now present a convenient strategy to estimate a quantile model with endogenous individual effects. A practical formulation for $g(\cdot)$ is to use a known parametric function of time-series averages or, alternatively, a vector of covariates for each of the N subjects. A one-step estimator is obtained by solving the following problem:

$$(2.6) \qquad \min_{\beta,\gamma,a \in \mathcal{B} \times \mathcal{G} \times \mathcal{A}} \sum_{j=1}^{J} \sum_{t=1}^{T} \sum_{i=1}^{N} \omega_j \rho_{\tau_j}(y_{it} - \boldsymbol{x}'_{it}\boldsymbol{\beta}(\tau_j) - \boldsymbol{x}'_i\boldsymbol{\gamma}(\tau_j) - a_i) + \lambda\|\boldsymbol{a}\|_1,$$

where as before ω_j is a relative weight given to the j-th quantile. It has been argued that the choice of the weights, ω_j, and the associated quantiles τ_j, is somewhat analogous to the choice of discretely weighted L-statistics (Koenker 2004). An alternative less efficient, yet practical choice, is to ignore the potential gains and estimate models with equal weights defined as $\omega_j = J^{-1}$ for all j.

2.2. Inference and selection of the tuning parameter

The solutions of (2.5) and (2.6) are a family of estimates in which each estimate is indexed by a parameter value of the tuning parameter, λ. Consider for example the argument that minimizes (2.6). As shown in Section 5.1, the family of estimates associated with the slope coefficients lies on a one-dimensional path of finite length in the $J(p+1)$-dimensional

space of slope coefficients simultaneously estimated. Our goal however is to reduce the computational burden and find a choice of λ, say λ^*, that is optimal with respect to a criterion function. The tuning parameter can be selected by a modified AIC-type approach, $\hat{\lambda} = \arg\inf \|\hat{u}(\tau, \lambda)\|_1 + \text{df}_\lambda/(2NT)$, where $\hat{u}(\tau, \lambda)$ is the residual and df_λ is the number of nonzero estimated parameters. Alternatively, the tuning parameter can be selected to minimize a quadratic loss function. Lamarche (2010) shows that the ℓ_1 penalty function $\|\boldsymbol{a}\|_1$ does achieve unbiasedness when the a_i's are drawn from zero-median distribution function, proposing to find, $\hat{\lambda} = \arg\inf \{tr\boldsymbol{\Sigma_\beta}\}$, where $\boldsymbol{\Sigma_\beta}$ is the covariance matrix of the slope parameter. The empirical covariance matrix can be easily obtained given λ and B bootstrap estimates $\{\hat{\boldsymbol{\beta}}^*(\boldsymbol{\tau}, \lambda), \hat{\boldsymbol{\gamma}}^*(\boldsymbol{\tau}, \lambda), \hat{\boldsymbol{a}}^*(\lambda)\}$. These bootstrap estimates are obtained sampling pairs $\{(\boldsymbol{y}_i, \boldsymbol{x}_i) : i = 1, ..., N\}$ with replacement.

3. Asymptotic Mean Squared Error

This section investigates the performance of ℓ_1 estimators for panel data under large N, large T asymptotics. We restrict the analysis to the one-step estimator under the regularity conditions stated in Koenker (2004) because they facilitate the comparison of the proposed method with existing ℓ_1-penalized methods. We compare the performance of three estimators: the estimator that penalizes uncorrelated individual effects $\hat{\boldsymbol{\beta}}(\tau, \lambda)$, the estimator that penalizes correlated individual effects $\tilde{\boldsymbol{\beta}}(\tau, \lambda)$, and the quantile regression estimator for the correlated random effects model $\hat{\boldsymbol{\beta}}(\tau, 0)$. The estimator $\tilde{\boldsymbol{\beta}}(\tau, \lambda)$ is similar to the estimator considered in Koenker (2004) when the location of the distribution of the *iid* α_i's is different than zero, and $\hat{\boldsymbol{\beta}}(\boldsymbol{\tau}, 0)$ is similar to the estimator considered in Abrevaya and Dahl (2008) replacing the time effects by individual effects.

The appendix presents the assumptions and definitions associated with the main results of this section. Nevertheless, we briefly introduce notation for convenience. Let $\boldsymbol{H}_1$, $\boldsymbol{\Sigma}_1$, $\boldsymbol{J}_0$, $\boldsymbol{J}_2$, $\boldsymbol{J}_3$ be limiting positive definite matrices, $\boldsymbol{L}$ is a weighted orthogonal projection matrix of independent variables $\boldsymbol{X}$, and $\boldsymbol{\Phi}$ and $\boldsymbol{\Upsilon}$ denote diagonal matrices. Moreover, define $\boldsymbol{A} = \boldsymbol{J}_3$, $\boldsymbol{C} = \boldsymbol{J}_2$, $\boldsymbol{D} = \boldsymbol{J}_2^{-1}\boldsymbol{J}_0$, $\tilde{\boldsymbol{B}} = \boldsymbol{J}_3^{-1}\boldsymbol{L}'\boldsymbol{\Phi}\boldsymbol{L}$ and $\hat{\boldsymbol{B}} = \boldsymbol{J}_3^{-1}\boldsymbol{L}'\boldsymbol{\Upsilon}\boldsymbol{L}$, and ζ_a, $\zeta_{\tilde{b}}$, $\zeta_{\hat{b}}$, ζ_c, ζ_d the corresponding positive eigenvalues of the matrices. Lastly, define $\boldsymbol{s}_{0,it} = (\mathbb{E}(\text{sign}(\alpha_i)\boldsymbol{x}_{it}))_{it}$ and $\boldsymbol{S}_0 = \boldsymbol{s}_0\boldsymbol{s}_0'$. The largest eigenvalue is defined as $\bar{\zeta}_{S_o} = \max\{\zeta_{S_o}^1, \ldots, \zeta_{S_o}^p\}$, and $\bar{\zeta}_{\tilde{a}}$ and $\bar{\zeta}_{\tilde{b}}$ are the corresponding eigenvalues.

THEOREM 1. *Under the conditions provided in the Appendix, for $\lambda \in (0, \infty)$, the penalized estimator that shrinks endogenous individual effects, $\tilde{\boldsymbol{\beta}}(\tau, \lambda)$, and the penalized estimator that*

shrinks exogenous individual effects, $\hat{\boldsymbol{\beta}}(\tau, \lambda)$, have covariance matrices,

$$Avar(\sqrt{NT}(\tilde{\boldsymbol{\beta}}(\tau, \lambda)) = (\boldsymbol{H}_1 + \lambda \boldsymbol{J}_3)^{-1}(\boldsymbol{J}_0 + \lambda^2 \boldsymbol{J}_2)(\boldsymbol{H}_1 + \lambda \boldsymbol{J}_3)^{-1},$$

$$Avar(\sqrt{NT}(\hat{\boldsymbol{\beta}}(\tau, \lambda)) = (\boldsymbol{\Sigma}_1 + \lambda \boldsymbol{J}_3)^{-1}(\boldsymbol{J}_0 + \lambda^2 \boldsymbol{J}_2)(\boldsymbol{\Sigma}_1 + \lambda \boldsymbol{J}_3)^{-1},$$

and $Avar(\tilde{\boldsymbol{\beta}}(\tau, \lambda)) < Avar(\hat{\boldsymbol{\beta}}(\tau, \lambda))$. Also,

$$|Abias(\tilde{\boldsymbol{\beta}}(\tau, \lambda))| > |Abias(\hat{\boldsymbol{\beta}}(\tau, \lambda))| = |Abias(\hat{\boldsymbol{\beta}}(\tau, 0))| = \boldsymbol{0}.$$

The result could be interpreted in terms of asymptotic mean squared error (AMSE). Note that although $\tilde{\boldsymbol{\beta}}(\tau, \lambda)$ is asymptotically biased, it may have asymptotically significantly smaller variance than the unbiased estimators $\hat{\boldsymbol{\beta}}(\tau, \lambda)$ and $\hat{\boldsymbol{\beta}}(\tau, 0)$.

COROLLARY 1. *Under the conditions of Theorem 1, for $\lambda \in (0, \infty)$, the trace of the asymptotic mean squared error of the penalized estimator that shrinks endogenous individual effects $\tilde{\boldsymbol{\beta}}(\tau, \lambda)$, and the penalized estimator that shrinks exogenous individual effects $\hat{\boldsymbol{\beta}}(\tau, \lambda)$ are:*

$$AMSE(\tilde{\boldsymbol{\beta}}(\tau, \lambda)) = \sum_{i=1}^{p} \frac{\zeta_c^i(\zeta_d^i + \lambda^2)}{(\zeta_a^i(\zeta_{\tilde{b}}^i + \lambda))^2} + \frac{\bar{\zeta}_{S_o}\lambda^2}{(\bar{\zeta}_a(\bar{\zeta}_{\tilde{b}} + \lambda))^2}$$

$$AMSE(\hat{\boldsymbol{\beta}}(\tau, \lambda)) = \sum_{i=1}^{p} \frac{\zeta_c^i(\zeta_d^i + \lambda^2)}{(\zeta_a^i(\zeta_{\hat{b}}^i + \lambda))^2}.$$

It is immediately apparent than for $\bar{\zeta}_{S_o}$ sufficiently small,

$$\text{AMSE}(\tilde{\boldsymbol{\beta}}(\tau, \lambda)) \leq \text{AMSE}(\hat{\boldsymbol{\beta}}(\tau, \lambda)),$$

because $\zeta_{\tilde{b}}^i > \zeta_{\hat{b}}^i$ for all i by Weyl's monotonicity principle of eigenvalues (Bhatia 1997), but the inequality is reversed if the bias and the tuning parameter are large. For λ sufficiently small, we have that $\text{AMSE}(\hat{\boldsymbol{\beta}}(\tau, \lambda)) \leq \text{AMSE}(\hat{\boldsymbol{\beta}}(\tau, 0))$, suggesting that shrinking the individual effects a's is worthwhile.

It seems natural to consider choosing λ to minimize AMSE, which for the case of $\hat{\boldsymbol{\beta}}(\boldsymbol{\tau}, \lambda)$ implies choosing λ to minimize asymptotic variance. The primary objective is now to show that the trace of the asymptotic covariance matrix of $\hat{\boldsymbol{\beta}}(\boldsymbol{\tau}, \lambda)$ is convex in λ, therefore a unique value of λ exists. In contrast, the variance of $\hat{a}_i(\lambda)$, which is not derived in Theorem 1, is expected to tend monotonically to zero as λ tends to infinity.

The following result demonstrates that it is possible to obtain an optimal tuning parameter defined as the minimizer of the trace of the asymptotic covariance matrix. Note that the

selection of λ^* is not sensitive to scale effects because we consider normalized asymptotic variances $\mathrm{Avar}(\hat{\beta}_k(\tau, \lambda))/\mathrm{Avar}(\hat{\beta}_k(\tau, 0))$.

COROLLARY 2. *Under the conditions of Theorem 1, there exists a unique variance minimizing parameter, $\lambda^* = \arg\min\{tr(\boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}_0^{-1} \boldsymbol{\Sigma}_1)(\boldsymbol{\Sigma}_1 + \lambda \boldsymbol{\Sigma}_3)^{-1}(\boldsymbol{\Sigma}_0 + \lambda^2 \boldsymbol{\Sigma}_2)(\boldsymbol{\Sigma}_1 + \lambda \boldsymbol{\Sigma}_3)^{-1}\}$.*

Standard arguments can be used to construct a "plug-in" estimator $\hat{\lambda}$ that consistently estimates the optimal degree of shrinkage λ^*. The estimation of the asymptotic covariance matrix can be accomplished by obtaining estimates of the conditional density f at the conditional quantile $\xi(\tau)$ and the density of the individual effects $g(0)$. The estimation of $f(\xi(\tau))$ in iid and non-iid settings requires the use of standard quantile regression methods, considering the conditional quantile function evaluated at λ equal to zero, $\xi(\tau, 0)$. The interested reader will find in Koenker (2005) detailed explanations on the existing approaches. In the case of location-scale shift models, the estimation of $g(0)$ can be performed considering a sample of normalized individual effects estimates $\{\hat{a}_1(0), \hat{a}_2(0), ..., \hat{a}_N(0)\}$ and classical kernel methods, $(1/(Nh_N)) \sum_{i=1}^{N} K(\hat{a}_i(0)/h_N)$, where h_N is a bandwidth and $\hat{a}_i(0)$ is the "unpunished" estimate of the individual effect a_i. More general models can be estimated using the bootstrap procedure described in Section 2.2.

4. Monte Carlo

This section reports the results of several simulation experiments designed to evaluate the performance of the method in finite samples. First, we will briefly investigate the bias and variance of the penalized estimator in models with endogenous individual effects. Second, we will contrast the performance of the penalized quantile regression estimator for the correlated random effects model with classical least squares estimators and quantile regression estimators. Finally, we will evaluate the efficiency of the penalized estimator relative to existing approaches for panel data quantile regression.

4.1. Experiment designs and methods

We generate the dependent variable considering the following version of the model (2.1)-(2.3):

$$
\begin{aligned}
y_{it} &= \beta_0 + \beta_1 x_{it} + \alpha_i + (1 + \delta x_{it})u_{it}, \\
x_{it} &= \pi \mu_i + v_{it}, \\
\alpha_i &= g(\gamma_0 + \gamma_1 x_{i1} + ... + \gamma_T x_{iT}) + a_i
\end{aligned}
$$

The first three designs consider the location shift model $\delta = 0$ and the last design assumes a location-scale shift model with $\delta = 1$:

Design 1: The function $g(\cdot)$ is assumed to be known and linear and u_{it} is $\mathcal{N}(0,1)$. The variables μ_i, v_{it}, and a_i are iid Gaussian variables. The parameter of interest β_1 is assumed to be zero, the γ's are $0.5/T$ representing the Mundlak-Chamberlain case, and π is set to be 2.5.

Design 2: The function $g(\cdot)$ is nonlinear. We assume that $g(\cdot) = \sin(\cdot)$ and $\gamma_t = 2\pi/T$ for all t. This implies that $\alpha_i = \sin(2\pi\bar{x}_i) + a_i$ where $\bar{x}_i$ denotes the individual-specific average of x_{it}. The variables μ_i, v_{it}, and a_i are iid Gaussian variables.

Design 3: We maintain Design 1 with the exception of the distribution of the error term u_{it} that is now distributed as t-student with three degrees of freedom, t_3.

Design 4: We reproduce the first design used in Canay (2011). The function $g(\cdot)$ is assumed to be linear and $\gamma_t = 2$ for all t. The variable $u_{it} \sim \mathcal{N}(2,1)$, $a_i \sim \mathcal{N}(0,1)$ and $v_{it} \sim Beta(1,1)$. The parameter $\beta_0 = -1$ and the parameters $\beta_1 = \pi = 0$.

In the next section, we employ several sample sizes $N = \{100, 500\}$ and $T = \{2, 5, 12\}$ and compare the performance of the following estimators: (1) the ordinary least squares (OLS); (2) the generalized least squares (GLS); (3) the pooled quantile regression estimator (QR); (4) Koenker's (2004) penalized quantile regression estimator for a model with fixed effects that uses the optimal tuning parameter proposed in Lamarche (2010) (PFE); (5) Abrevaya and Dahl's (2008) quantile regression estimator for the correlated random effects model (CQR); (6) Canay's (2011) two-step fixed effects quantile regression estimator (2SQR); (7) the semiparametric penalized quantile regression estimator (SQR) defined in equation (2.5); and (8) two penalized quantile regression estimators for the linear parametric correlated random effects model (PQRd and PQR), defined in equation (2.6). The estimator labelled PQRd allows $\gamma(\tau)$'s to vary by quantile, while PQR assumes that $\gamma_t(\tau) = \gamma_t$ for all τ. The empirical evidence is based on 400 samples.

4.2. Results

We start reporting results on the performance of the penalized quantile regression estimator in parametric and semiparametric models. Considering $N = 100$ and $T = 5$, Figure 4.1 report the bias and variance percentage change of PFE, PQR and SQR. The upper panels present evidence of the performance of these three methods when the data is generated according to Design 1 and the lower panels present evidence when the data is generated as in Design 2. We see that the upper left panel shows that the PFE estimator is biased, and
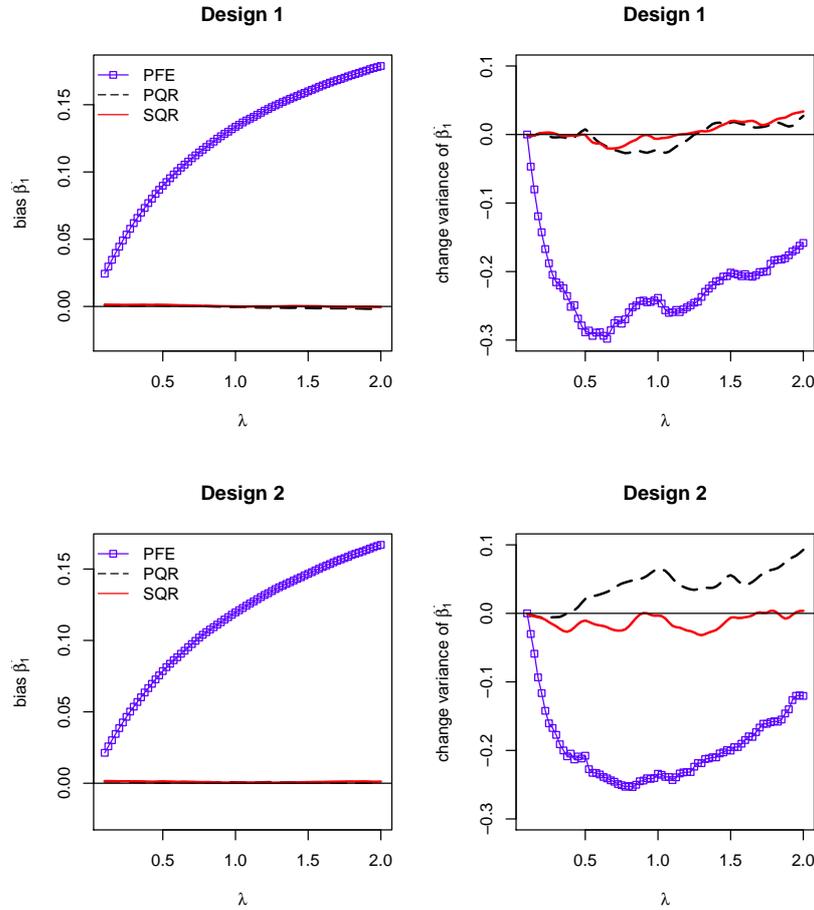
FIGURE 4.1. *Small sample performance of penalized quantile regression estimator for $\beta_1(0.5)$. The left panel shows the bias of the estimator and the right panel shows the variance percentage change.*

its bias starts to increase as we increase the harshness of the penalization. The right upper panel reveals that (i) the variance of the estimator decreases first and then increases, and (ii) there are significant differences in variance reduction. By carefully choosing λ to be 1, the variance of the slope PQR estimator is reduced more than 30 percent, while the variance of the PQR is reduced 2 percent.[1]

The PQR estimator is the efficient estimator in the class of penalized estimators for panel data. As expected however the performance of this method is rather unsatisfactory in terms

---

[1]Additional evidence not reproduced here to save space showed that the variances of the estimators are not influenced by the correlation between α_i and x_{it}, but as expected, the bias of the PFE estimator does.

N	T		Least Squares			Quantile Regression					
								Estimators			
			OLS	GLS	QR	PFE	CQR	2SQR	SQR	PQRd	PQR
						Design 1: $\tau = 0.50$					
100	2	Bias	0.458	0.390	0.457	0.407	-0.008	-0.011	-0.010	-0.009	-0.010
		RMSE	0.460	0.392	0.460	0.410	0.140	0.096	0.106	0.115	0.107
100	12	Bias	0.424	0.150	0.423	0.140	0.002	0.002	0.002	0.002	0.002
		RMSE	0.425	0.152	0.425	0.143	0.042	0.032	0.036	0.039	0.035
500	2	Bias	0.464	0.408	0.465	0.421	0.005	0.001	0.000	0.002	0.001
		RMSE	0.465	0.408	0.465	0.422	0.063	0.041	0.047	0.053	0.047
500	12	Bias	0.440	0.180	0.441	0.165	-0.001	0.000	0.000	0.000	0.000
		RMSE	0.441	0.180	0.441	0.166	0.018	0.014	0.015	0.016	0.015
						Design 3: $\tau = 0.50$					
100	2	Bias	0.461	0.393	0.460	0.411	0.006	-0.007	0.000	0.001	0.001
		RMSE	0.468	0.401	0.466	0.417	0.174	0.155	0.136	0.139	0.134
100	12	Bias	0.424	0.148	0.421	0.140	-0.001	-0.001	-0.001	-0.002	-0.001
		RMSE	0.430	0.154	0.424	0.145	0.053	0.049	0.039	0.042	0.039
500	2	Bias	0.464	0.408	0.464	0.423	0.002	0.004	0.002	0.001	0.003
		RMSE	0.466	0.409	0.465	0.424	0.073	0.070	0.057	0.063	0.058
500	12	Bias	0.439	0.179	0.439	0.170	-0.001	0.000	0.000	0.000	0.000
		RMSE	0.439	0.180	0.439	0.171	0.024	0.022	0.018	0.020	0.018
						Design 4: $\tau = 0.50$					
100	2	Bias	0.827	0.375	0.845	0.411	-0.021	-0.006	-0.013	-0.006	0.000
		RMSE	1.761	0.894	1.854	1.001	0.779	0.537	0.591	0.590	0.572
100	12	Bias	0.818	0.071	0.818	0.050	-0.004	0.005	0.007	0.002	0.004
		RMSE	1.651	0.211	1.669	0.217	0.238	0.189	0.190	0.192	0.190
500	2	Bias	1.011	0.505	1.011	0.518	-0.005	-0.013	-0.013	-0.012	-0.013
		RMSE	2.040	1.032	2.053	1.067	0.352	0.243	0.265	0.273	0.262
500	12	Bias	0.924	0.075	0.934	0.045	0.001	-0.002	-0.002	-0.003	-0.002
		RMSE	1.851	0.166	1.875	0.129	0.114	0.089	0.092	0.093	0.092
						Design 4: $\tau = 0.25$					
100	2	Bias	1.757	1.075	1.440	0.674	0.161	0.336	0.135	0.166	0.154
		RMSE	2.405	1.505	2.098	1.114	0.943	0.709	0.718	0.820	0.743
100	12	Bias	1.743	0.615	1.537	0.097	0.213	0.069	0.029	0.029	0.028
		RMSE	2.321	0.831	2.077	0.249	0.390	0.218	0.214	0.209	0.215
500	2	Bias	2.034	1.271	1.738	0.840	0.176	0.316	0.143	0.099	0.140
		RMSE	2.710	1.698	2.336	1.152	0.456	0.490	0.366	0.389	0.359
500	12	Bias	1.902	0.621	1.647	0.094	0.203	0.062	0.021	0.021	0.020
		RMSE	2.524	0.827	2.191	0.160	0.296	0.126	0.104	0.106	0.104

TABLE 4.1. *Small sample performance of the methods.*
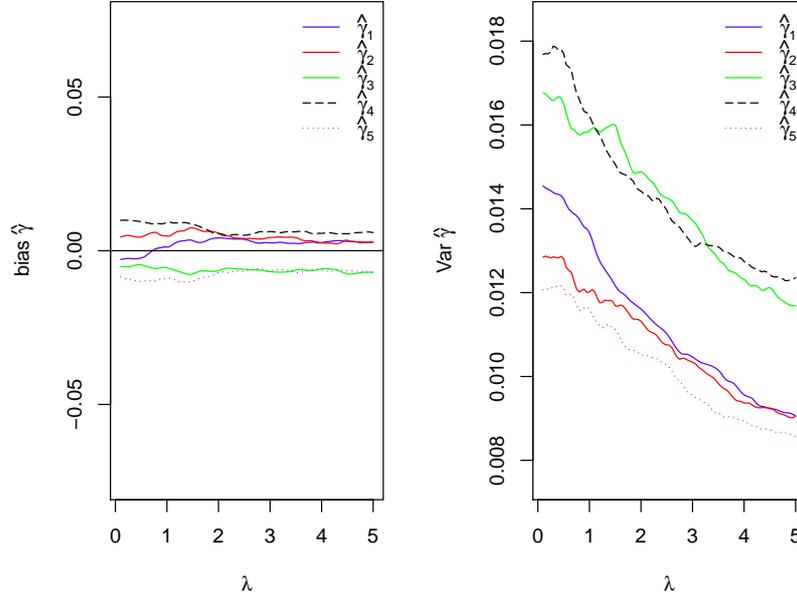
FIGURE 4.2. *Small sample performance of the penalized quantile regression estimator PQR for* $\boldsymbol{\gamma}$.

of bias in models with endogenous individual effects. In contrast, PQR and SQR are unbiased. It is interesting to see that the performance of the PQR estimator deteriorates when $g(\cdot)$ is a non-linear function but the performance of the semi-parametric version of the estimator remains essentially the same.

Figure 4.2 reports the bias and variance of the estimator of $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, ..., \gamma_5)'$ under Design 1. The figure shows that the bias of $\hat{\gamma}_t$ is small for all t and λ. The figure also shows that the variance tends to decrease as when increase λ reflecting potential significant improvements in the performance of the estimator that penalizes exogenous individual effects.

Using Table 4.1, we expand the design of the experiment considering several sample sizes $N = \{100, 500\}$ and $T = \{2, 12\}$, and the random variables a_i and u_{it} to be distributed as Gaussian and t-student with 3 degrees of freedom. The table also reports evidence of the performance of the methods under Design 4. As expected, the performances of the methods that ignore the correlation between the independent variable and the individual effect are rather unsatisfactory. In all the variants of the model, the bias is significant even for moderate T. Notice also that the estimators that address the endogeneity of the α_i's

		Asymptotic Theory				Bootstrap			
		Design 1		Design 3		Design 1		Design 3	
N	T	0.25	0.50	0.25	0.50	0.25	0.50	0.25	0.50
100	2	0.1467	0.1142	0.1872	0.1429	0.1414	0.1258	0.1795	0.1499
		[0.9329]	[0.7922]	[0.8604]	[0.8185]	[0.8991]	[0.8722]	[0.8253]	[0.8585]
100	12	0.0416	0.0377	0.0497	0.0407	0.0409	0.0384	0.0500	0.0412
		[0.8437]	[0.8644]	[0.7947]	[0.8038]	[0.8292]	[0.8794]	[0.8056]	[0.8020]
500	2	0.0656	0.0520	0.0865	0.0655	0.0629	0.0544	0.0834	0.0666
		[0.9662]	[0.8378]	[0.9838]	[0.8774]	[0.9268]	[0.8770]	[0.9493]	[0.8912]
500	12	0.0193	0.0169	0.0229	0.0192	0.0195	0.0167	0.0236	0.0195
		[0.8862]	[0.8646]	[0.7753]	[0.8504]	[0.8980]	[0.8572]	[0.8017]	[0.8642]

TABLE 4.2. *Standard error and relative efficiency of CQR relative to PQR (in brackets).*

produce unbiased results. In the location-scale shift model at $\tau = 0.25$, PQR and SQR estimators offer the best overall performance related to existing methods.

The results presented in Table 4.1 might be sensitive to selecting the tuning parameter $\lambda = \lambda^*$. In practice, the optimal degree of shrinkage is not known and therefore it has to be estimated. We now briefly investigate the performance of PQR estimator considering two ways of estimating λ^*: (i) estimated asymptotic covariance matrix and (ii) bootstrapped variance. Table 4.2 suggests that there are no important efficiency losses when the practitioner estimates λ^*, at least in the models considered in this study. Second, the performance of the two λ selection alternatives are satisfactory. Lastly, the PQR estimator seems to advance the CQR estimator, offering considerable efficiency gains in all variants of the model.

5. Applications to models with heterogeneous preferences

This section illustrates the use of the panel data quantile regression approach that handles the presence of potentially endogenous α's in applications with heterogeneous preferences.

5.1. New evidence on the responsiveness of hours to wages

We first show that the empirical model could be motivated by a simple variation of the neoclassical labor-supply model. Considering a sample from the British Household Panel Survey (BHPS), we demonstrate that the approach produces similar results to the quantile regression version of the fixed effects estimator, but it allows for identification of effects associated with time-invariant taste shifters and demographic characteristics. Additionally,

it allows us to estimate wealth effects of a parametric change in wages on hours worked. We find evidence that positive substitution effects dominate negative wealth effects at several quantiles of the conditional distribution of hours. The overall effect tends to be decreasing across quantiles, which might be interpreted as suggesting that changes in taxes could have different effects across the distribution of hours.

5.1.1. *Model specification*

Before turning to the discussion of the results, we briefly revisit the traditional life cycle labor supply model (see, e.g., Blundell and MaCurdy 1999, Pistaferri 2003). It is standard to write,

$$\ln(h) \approx \boldsymbol{x}'\boldsymbol{\beta} + \delta \ln(w) + \pi \ln(\alpha), \tag{5.1}$$

where the parameter δ denotes the intertemporal substitution elasticity of labor supply and the term $\pi \ln(\alpha)$ denotes the marginal utility of wealth. This simple empirical labor supply model (5.1) is motivated by the convenient additive separable utility function $U(c, h; v) = c - \exp(\boldsymbol{x}'\boldsymbol{\beta})h^{1+v}$, where c is consumption and $v > 0$ is a time-invariant, consumer-invariant parameter. To obtain a life-cycle specification, one may assume that the marginal utility of wealth is,

$$\pi \ln(\alpha_{i0}) = \boldsymbol{z}'_{i0}\boldsymbol{\theta} + \sum_{s=0}^{T} \gamma_s \mathbb{E}_0\{\ln w_{is}\} + a_i, \tag{5.2}$$

where $\boldsymbol{z}$ is a vector of demographic characteristics that may include non-labor income, and, by definition, the individual idiosyncratic term a_i is orthogonal to the covariates.[2] Notice that for estimating (5.1)-(5.2), the practitioner requires information on consumers' expected wages.

We introduce a simple variation of the labor supply function that leads to a more flexible empirical framework. We simply assume that the consumer solves the maximization of the

---

[2]Equation (5.2) assumes that each consumer knows her working life of T_i periods. The main empirical concern of this theoretical restriction is that wages outside the sample period are not observed. We try to accommodate this by considering the largest possible panel (e.g., we drop consumers for which we have T_i less than 12 years). Another standard limitation is associated with measuring initial wealth. We will estimate models for workers with and without investments at time 0, to evaluate the sensitivity of our main conclusions. Alternatively, it is possible to assume quadratic equations in age for approximating the profiles of wages and property income. For instance, $\mathbb{E}_0\{\ln w_{is}\} = \pi_0 + \pi_1 s + \pi_2 s^2 + u_{is}$ (see, e.g., Blundell and MaCurdy 1999).

lifetime utility function $U(c, h; v(u))$, subject to an intertemporal budget constraint. The function $v(u)$ describes heterogeneity in preferences. Similar to before, assuming an interior optimum, it is possible to obtain that,

$$(5.3) \qquad \ln(h) \approx \boldsymbol{x}'\boldsymbol{\beta}(u) + \delta(u)\ln(w) + \pi(u)\ln(\alpha),$$

where $\delta(u) = v(u)^{-1}$, $\boldsymbol{\beta}(u) = \boldsymbol{\beta}/v(u)$, and $\pi(u) = \pi/v(u)$. This simple way of introducing heterogeneity in preferences does not require one to assume that the distribution function for latent heterogeneity is known. Additionally, it relaxes the assumption that δ represents a simple location-shift on the conditional distribution of hours. Therefore, we write,

$$(5.4) \qquad \pi(u)\ln(\alpha_{i0}) = \boldsymbol{z}'_{i0}\boldsymbol{\theta}(u) + \sum_{s=1}^{T} \gamma_s(u)\ln w_{is} + a_i,$$

The model presented in equations (5.3) and (5.4) has the following random coefficient representation,

$$(5.5) \qquad \ln(h_{it}) = \boldsymbol{c}'_{it}\boldsymbol{\psi}(u_{it}) + (\delta(u_{it}) + \gamma_t(u_{it}))\ln w_{it} + \sum_{s\neq t} \gamma_s(u_{it})\ln w_{is}$$

$$(5.6) \qquad \tau \mapsto \boldsymbol{c}'_{it}\boldsymbol{\psi}(\tau) + (\delta(\tau) + \gamma_t(\tau))\ln w_{it} + \sum_{s\neq t} \gamma_s(\tau)\ln w_{is}$$

where unmeasured characteristics $u_{it}|\boldsymbol{c}_{it}, \ln\boldsymbol{w}_i$ are uniformly distributed $\mathcal{U}(0,1)$, and τ denotes the τ-th quantile of the conditional distribution of the logarithm of hours. The variable $\boldsymbol{c}_{it} = (\boldsymbol{x}'_{it}, \boldsymbol{z}'_i, d_{it})'$ includes taste shifters, demographic characteristics and an indicator variable d_{it} for the individual effect a_i, and $\boldsymbol{\psi}(\tau) = (\boldsymbol{\beta}(\tau)', \boldsymbol{\theta}(\tau)', \boldsymbol{a}')'$.[3]

We estimate the panel data-quantile model (5.6) considering the method proposed above. We restrict attention to the elasticity of labor supply with respect to a parametric change in wages, which is a combination of substitution and wealth effects. We define the term $\Gamma(\tau) = \sum_t \gamma_t(\tau)$ as the wealth effect of a parametric shift in the wage *profile* at time 0, and $\delta(\tau)$ as the substitution elasticity at the τ-th quantile of the conditional distribution of hours. The elasticities are quantile-specific allowing us to go beyond the standard wage elasticity for the "representative" consumer.

---

[3]Notice that MaCurdy's (1981) representation of the endogenous individual effects lies within the correlated random effects framework, if time-variant variables excluding wages do not affect the marginal utility of wealth. In terms of the model estimated in this section, this suggests that age should not be considered in (5.4).

Variables	All workers	Workers with zero non-labor income	Workers with positive non-labor income
	(1)	(2)	(3)
Hours worked	36.749	38.650	36.232
	(10.511)	(13.044)	(9.646)
Wage rate (in logs)	5.738	5.611	5.773
	(0.679)	(0.696)	(0.670)
Age	39.403	37.456	39.933
	(9.008)	(8.583)	(9.049)
Union member	0.634	0.651	0.629
	(0.482)	(0.477)	(0.483)
Male worker	0.500	0.542	0.488
	(0.500)	(0.498)	(0.500)
Intermediate qualifications	0.425	0.530	0.325
	(0.494)	(0.499)	(0.469)
Advanced qualifications	0.490	0.325	0.534
	(0.500)	(0.469)	(0.499)
Number of workers	388	83	305
Number of observations	4656	996	3660

TABLE 5.1. *Descriptive statistics for the BHPS sample. The table presents the sample mean and standard deviation (in parenthesis)*

5.1.2. *Data description*

We use a sample of workers taken from the British Household Panel Survey (BHPS), which is similar to other data used in previous labor supply studies (e.g., Panel Study of Income Dynamics in the US). The sample includes $N = 388$ and $T = 12$ (Table 5.1). The period of analysis is 1991-2002. The data contains observations on weekly hours worked, real wages, demographic and socioeconomic characteristics including union membership in 1991, and gender.[4]

We also include variables for the level of education of the worker, deviating from the traditional continuous years to schooling measure (Card 1999). Recently, Battistin and Sianesi

---

[4]The empirical literature, in general, and studies related with the BHPS, in particular, distinguish the estimation of labor supply by gender. Although it is obviously possible to estimate the model presented in (5.6) with a subsample of male workers, we include a variable indicating gender in the quantile regression model. Addressing issues related to whether the individual works in the labor market is out of the scope of this paper.

| | Quantiles | | | | | | | | |
| | 0.10 | | | 0.50 | | | 0.90 | | |
	QR	FEQR	PQR	QR	FEQR	PQR	QR	FEQR	PQR
Substitution	0.427	0.421	0.443	0.225	0.339	0.290	0.130	0.284	0.296
elasticity (δ)	(0.036)	(0.044)	(0.048)	(0.039)	(0.041)	(0.036)	(0.032)	(0.044)	(0.044)
Union	0.027		-0.018	-0.008		-0.008	-0.031		-0.003
	(0.033)		(0.028)	(0.020)		(0.023)	(0.029)		(0.024)
Male	0.143		0.110	0.103		0.107	0.209		0.097
	(0.042)		(0.035)	(0.016)		(0.025)	(0.030)		(0.029)
Intermediate	-0.046		-0.094	-0.074		-0.067	-0.047		-0.052
qualifications	(0.060)		(0.039)	(0.028)		(0.037)	(0.054)		(0.038)
Advanced	-0.274		-0.203	-0.206		-0.175	-0.139		-0.149
qualifications	(0.070)		(0.055)	(0.036)		(0.047)	(0.055)		(0.048)
Wealth effect			-0.176			-0.115			-0.140
(Γ)			(0.069)			(0.045)			(0.048)
Wage elasticity			0.267			0.175			0.156
($\delta + \Gamma$)			(0.053)			(0.047)			(0.048)

TABLE 5.2. *Estimating a panel quantile labor supply model. QR refers to quantile regression, FEQR is fixed effects quantile regression and PQR is the penalized estimator for the linear correlated random effects model. The value of the tuning parameter is 0.65.*

(2011) argue that in UK the return to education does not increase linearly with each additional year of schooling, because there are other academic paths leading to different qualifications. To avoid biases introduced by misclassification, we construct indicators for whether the worker has intermediate or advanced qualifications. The omitted category in the regressions is 'no qualifications'. While the level 'no qualifications' reflects dropping out of school, 'intermediate' qualifications includes level 2 - O levels or their vocational equivalent. The variable 'advanced' qualification is an indicator for high-school diploma, advanced and higher City and Guilds and Royal Society of Arts, Ordinary and Higher National Diplomas, university and postgraduate studies, and professional degrees.

5.1.3. *Estimating quantile specific elasticities*

We use Table 5.2 to compare results obtained from our method with standard quantile regression methods. The PQR estimator uses a plug-in estimate $\hat{\lambda} = 0.65$ selected as discussed in Section 3. We report results for the 0.10, 0.50 and 0.90 quantiles. The table presents estimates of the elasticity parameter and the time invariant effects as a function of the quantiles of the conditional distribution of hours.

In this application, the classical fixed effects method (not reported in Table 5.2 to save space) suggests a positive substitution elasticity between hours and wages of 0.43 percent. However, the quantile approaches seem to provide additional information suggesting that the location-shift model is inappropriate for the British data set. For instance, the quantile regression version of the fixed effects estimator gives positive estimates ranging from 0.42 percent at the 0.1 quantile to 0.28 percent at the 0.9 quantile. Although the fixed effects estimates might be unbiased, the table shows that time invariant effects are not identified. In contrast, the penalized approach offers the possibility of estimating the effect of union membership, gender, and educational attainment on the conditional distribution of hours worked. While the effect of union membership is insignificant across the quantiles of the conditional distribution, gender and the indicators for educational attainment are significant at standard levels.

The PQR approach offers the possibility of estimating the main parameter of interest for policy analysis, $\delta + \Gamma$. At first glance, we see that the penalized estimates of δ are similar to the fixed effects estimates. These estimates δ are positive, and they tend to decrease as we increase the quantile τ. The estimated wealth effect is negative as expected, and tends to be significant if we consider quantiles at the tails. Although the sign of the parameter $(\delta + \Gamma)$ is unknown, we find evidence that suggests that the substitution effect dominates the wealth effect among British workers. The implication is that a 10% increase in the wage profile increases labor supply by approximately 2.7% at the 0.1 quantile and 1.6% at the 0.9 quantile.[5]

The results presented in Table 5.2 suggests that the wage elasticity of labor supply is positive across quantiles, because the positive substitution effect dominates the negative wealth effect. We briefly investigate the sensitivity of this result to the choice of λ. At the top of Figure 5.1, we plot estimates of the substitution effect, the wealth effect, and the overall effect at 25 equally spaced λ's in the interval $[0, 2]$. While we see positive estimates for the elasticity of substitution, we find negative estimates for the wealth effect. This evidence indicates that the sign of the estimated elasticity $\hat{\delta} + \hat{\Gamma}$ does not appear to be related to the optimal value of the tuning parameter. Notice that the point estimates do not seem to change dramatically

---

[5]Pistaferri (2003) finds that the conditional mean estimate of the substitution elasticity is 0.70 and the conditional mean estimate of the wealth effect is -0.20. He uses a panel from the Bank of Italy Survey of Household Income and Wealth and subjective expectations formed at the consumer level of aggregation. Notice that the availability of subjective expectations data would make possible to estimate (5.2). Although this approach has several important advantages, it requires the use of information that typically the practitioner does not have.
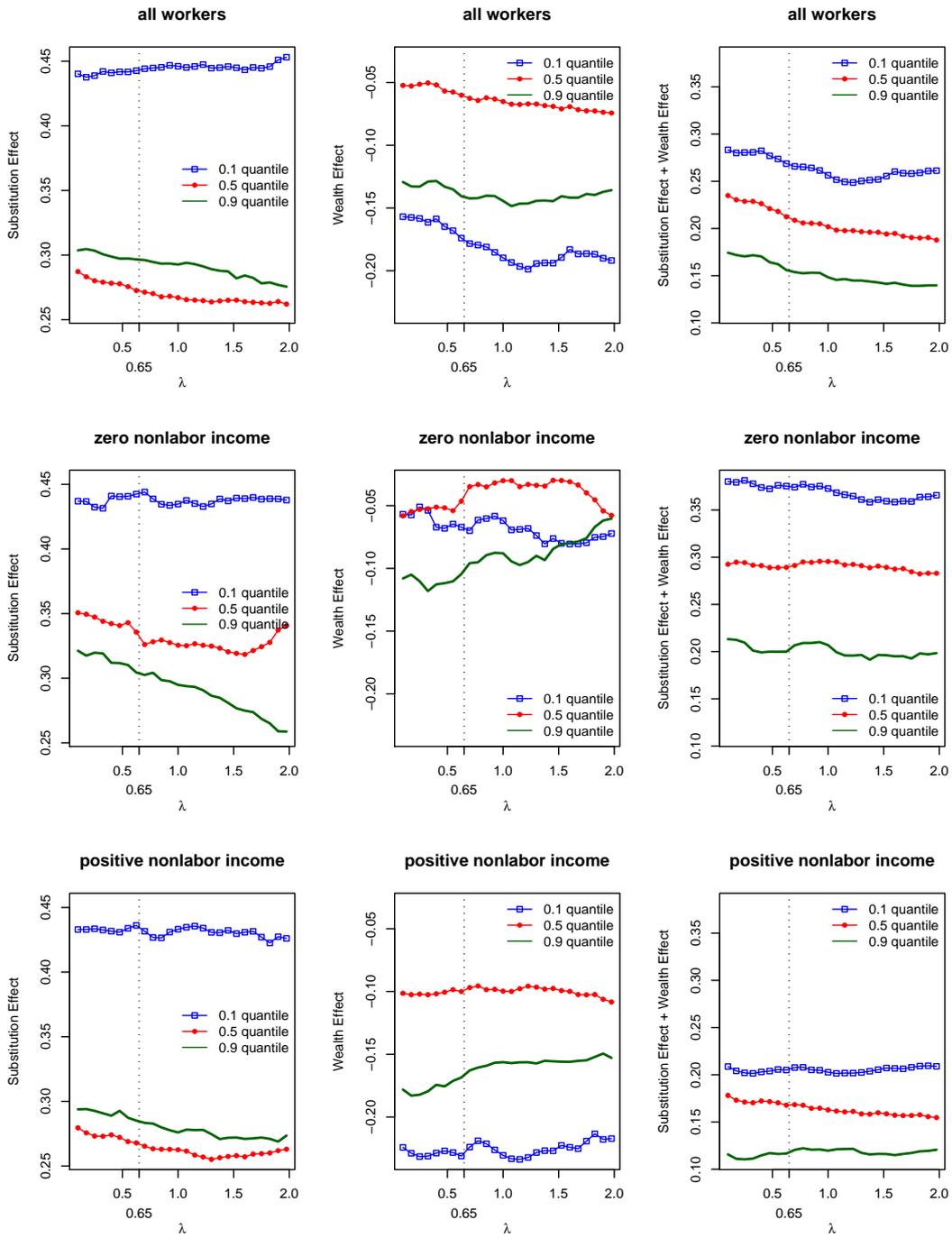
FIGURE 5.1. *Profile of the elasticities in terms of* λ.

as we increase λ, suggesting that the correlated effects framework provides a convenient way to account for individual heterogeneity associated with the covariates.

5.1.4. *Non-labor income and hours*

This section briefly investigates how asset income affects the allocation of hours. Based on the information provided by the BHPS, we define non-labor income as income from rents, savings and investment in the last year. We then re-estimate the model 5.6 splitting the sample in two subsamples. While the middle panels in Figure 5.1 present results based on a sample of workers with zero non-labor income, the panels at the bottom of the figure present results based on a sample of workers with positive non-labor income[6]. We note that the levels and shape of the estimated substitution effects do not change dramatically across samples, but the estimated wealth effects tend to be different across samples. While the wealth effects are close to zero and insignificant at standard levels among workers with zero non-labor income, the wealth effects are relatively large and significant among workers with positive non-labor income. Therefore, the evidence suggests that workers with zero non-labor income tend to be more sensitive to changes in wages than workers with positive non-labor income. Figure 5.1 also shows that the quantile specific elasticities provide a more informative analysis. The evidence could be interpreted as suggesting that while underemployed workers may be willing to offer more hours than the average worker given a one percent increase in wages, overemployed workers would like to offer fewer hours than the average worker.

5.2. **Heterogeneous nutrition preferences**

In the US obesity rates have increased at alarming rates over the last few decades. Given the comorbity of obesity with other chronic illnesses, such as diabetes and heart disease, and the financial strain it imposes on the health care system, obesity is considered to be one of the main public health concerns of our time. Numerous programs such as the First Lady's "Let's Move" campaign aim to address the challenge of obesity. Designing policies to address obesity is complex problem that requires us to tailor interventions to account for heterogeneous preferences and demographics.

---

[6]The second sample labeled 'zero non-labor income' considers workers without income from rents, savings and investments; the last sample labeled 'positive non-labor income' include workers with nonzero income from rents, savings, and investments.

In this example we explore how preferences for nutrients vary by gender. In a recent report the CDC[7] estimates that while 35.7% of US adults are now obese, substantial differences exist across genders and age groups. Over the last decade obesity rates for men increased from 27.5% to 35.5%, while obesity rates for women have not varied significantly in recent years and are currently at 35.8%. This section applies our method to quantifying preferences for nutrition by gender. Understanding the heterogeneity in preferences is of major policy interest as it helps us design better policies by accounting for their distributional impact. In the US context and in the face of rising obesity a variety of policies such as food taxes/subsidies, changes in the Supplemental Nutrition Assistance Program (SNAP), or portion restrictions are being considered, all of which will change the nutritional composition of a household's food consumption basket.

The consumption model described below allows us to quantify the differences in preferences for nutrients across sub-populations. Dubois, Griffith, and Nevo (2013) use this framework to show that there are substantial differences across countries. We adapt this model to explore if differences in preferences between socio-demographic groups exist at different levels of expenditure. We model a household's food purchase decision, where each household can choose among N different food products. Each product k is characterized by a set of D product attributes. Each product is identified at the UPC level. A typical American supermarket sells about $50,000$ different products. Thus, while D is large, the number of product attributes C which are salient to the average consumer is typically very small. We focus on attributes which relate to the underlying nutrients that a consumer may look for on the product label, such as calories, fat, salt, sugar, cholesterol, protein and carbohydrates. Each unit of good k contains $\varkappa_k = \{\varkappa_{k,1}, \ldots, \varkappa_{kC}\}'$ units of the underlying product attributes.[8]

We assume that household i with income I_i chooses a bundle of goods $\boldsymbol{x}_i$ and a numeraire good x_{i0} so as to maximize utility conditional on household attributes μ_i and subject to a budget constraint. We follow Dubois, Griffith, and Nevo (2013) and assume that the household derives utility from both the goods purchased and the underlying quantities $\boldsymbol{z}_i$ of the nutrients purchased through the purchase of the goods $\boldsymbol{x}_i$. We normalize the price of the numeraire good to 1. Denote by $\boldsymbol{p}_i$ the prices faced by household i over the set of products

---

[7]See Ogden, Carroll, Kit, and Flegal (2012) for more details.

[8]Note that we do not require the underlying product attributes to be mutually exclusive. For example Peanut Butter has 188 total calories per serving; 145 of those calories come from fat. The total amount of fat in the serving is 16.1 grams.

available for purchase. Thus,

$$(5.7) \qquad \max_{x_{i0}, \boldsymbol{x}_i} U(x_{i0}, \boldsymbol{x}_i, \boldsymbol{z}_i, \mu_i), \quad \text{s.t.} \quad x_{i0} + \boldsymbol{p}_i' \boldsymbol{x}_i = I_i,$$

where $z_{ik} = \varkappa_k' \boldsymbol{x}_i$, denotes a home production function which converts products into nutrients (cooking). This model generalizes the Muellbauer (1974) model of household production, which assumes that household i purchases goods $\boldsymbol{x}_i$ but only derives utility from the product attributes $\boldsymbol{z}_i$, and which generates a standard hedonic model. By allowing the household to derive utility from both products and attributes we are relying on insights from the modern Industrial Organization literature, which shows that households exhibit preferences over products.

Given the large number of choices D faced by the typical household, it would be impractical to estimate a disaggregated demand system. It is thus common to aggregate products into mutually exclusive groups such as milk products, meat products, etc. Harding and Lovenheim (2013) then estimate a structural Quadratic Almost Ideal Demand System on the aggregate set of products, which allows for the identification of a detailed substitution matrix. In the context of the current application, we wish to estimate the preference heterogeneity in the demand for nutrients and follow the more parametric approach of Dubois, Griffith and Nevo (2013) by choosing a utility which imposes stronger restrictions on the pattern of substitution between products. While this substantially limits the nature of the price elasticities, it possesses the attractive feature of weak separability, which leads to a convenient aggregation over products as shown below.

In particular we let:

$$(5.8) \qquad U(x_{i0}, \boldsymbol{x}_i, \boldsymbol{z}_i, \mu_i) = \exp(x_{i0}) \left(\sum_{k=1}^{D} f_{ik}(x_{ik}) \right)^{\mu_i} \prod_{c=1}^{C} h_{ic}(z_{ic}),$$

where μ_i denotes the set of household specific model parameters. We further assume that $f_{ik}(x_{ik}) = \lambda_{ik} x_{ik}^{\theta_i}$ and $h_{ic}(z_{ic}) = \exp(\beta_c z_{ic})$. After substituting for the budget constraint and the home production function the household maximizes the following log utility function:

$$(5.9) \quad \log U(x_{i0}, \boldsymbol{x}_i, \boldsymbol{z}_i, \mu_i, \lambda_i, \theta_i, \beta_c) = I_i - \sum_{k=1}^{D} p_{ik} x_{ik} + \mu_i \log \left(\sum_{k=1}^{D} \lambda_{ik} x_{ik}^{\theta_i} \right) + \sum_{c=1}^{C} \beta_c \sum_{k=1}^{D} \varkappa_{kc} x_{ik},$$

where $\varkappa_{kc}$ are known and observed by both the household and the econometrician. The parameters β_c measure the average contribution of a nutrient to the utility function. The

first order condition for good k is given by:

$$(5.10) \qquad p_{ik} = \mu_i \theta_i \frac{\lambda_{ik} x_{ik}^{\theta_i - 1}}{\sum_{k=1}^{D} \lambda_{ik} x_{ik}^{\theta_i}} + \sum_{c=1}^{C} \beta_c \varkappa_{kc}.$$

We can express this first order condition in terms of the expenditure for good k to obtain:

$$(5.11) \qquad p_{ik} x_{ik} = \mu_i \theta_i \frac{\lambda_{ik} x_{ik}^{\theta_i}}{\sum_{k=1}^{N} \lambda_{ik} x_{ik}^{\theta_i}} + \sum_{c=1}^{C} \beta_c \varkappa_{kc} x_{ik}.$$

Note that we can now aggregate this expression over all (or a subset) of the products to obtain relationship between total expenditure E_i and nutrients:

$$(5.12) \qquad E_i = \mu_i \theta_i + \sum_{c=1}^{C} \beta_c z_{ic}.$$

In practice we observe a household making repeated purchases and we can estimate this model using the following panel data empirical specification:

$$(5.13) \qquad E_{it} = \alpha_i + \sum_{c=1}^{C} \beta_c z_{itc} + \epsilon_{it},$$

where γ_t is a time indicator capturing the impact of macroeconomic factors on food purchases and ϵ_{it} is an iid error term uncorrelated with the RHS variables, capturing the random variation in consumer preferences. Furthermore we assume that the household specific effect α_i can be written as follows:

$$(5.14) \qquad \alpha_i = g(\boldsymbol{z}_i) + \boldsymbol{w}_i' \boldsymbol{\delta} + a_i,$$

where $g(\cdot)$ is an unknown function capturing household specific nutrition effects, $\boldsymbol{w}_i$ is a vector of time invariant household demographics and a_i is a household specific random effect, uncorrelated with the RHS variables.

We estimate this model using the following quantile representation,

$$(5.15) \qquad E_{it} \;\; = \;\; \sum_{c=1}^{C} \beta_c(u_{it}) z_{itc} + g(\boldsymbol{z}_i, u_{it}) + \boldsymbol{w}_i' \boldsymbol{\delta}(u_{it}) + a_i$$

$$(5.16) \qquad \tau \;\; \mapsto \;\; \sum_{c=1}^{C} \beta_c(\tau) z_{itc} + g(\boldsymbol{z}_i, \tau) + \boldsymbol{w}_i' \boldsymbol{\delta}(\tau) + a_i$$

where unmeasured characteristics conditional on observables are uniformly distributed $\mathcal{U}(0,1)$, and τ denotes the τ-th quantile of the conditional distribution of food expenditure.

Variables	All consumers	Female consumers	Male consumers
	(1)	(2)	(3)
Total Expenditure	120.328	119.887	121.311
	(77.954)	(77.558)	(78.821)
Total fat (grams consumed per month)	17.406	17.383	17.456
	(14.338)	(14.297)	(14.429)
Salt (miligrams consumed per month)	818.066	822.313	808.605
	(1173.582)	(1233.129)	(1028.525)
Sugar (grams consumed per month)	30.111	30.200	29.915
	(23.609)	(23.916)	(22.910)
Cholesterol (grams consumed per month)	46.541	46.705	46.176
	(42.987)	(42.764)	(43.479)
Protein (grams consumed per month)	13.852	13.668	14.262
	(23.351)	(24.363)	(20.914)
Carbohydrates (grams consumed per month)	64.023	64.068	63.924
	(46.093)	(46.989)	(44.032)
Black	0.117	0.128	0.093
	(0.322)	(0.334)	(0.291)
Income higher than $70K	0.146	0.123	0.197
	(0.353)	(0.329)	(0.398)
College education	0.475	0.450	0.530
	(0.499)	(0.497)	(0.499)
Number of consumers	9,165	6,326	2,839
Number of observations	109,980	75,912	34,068

TABLE 5.3. *Descriptive statistics for the Nielsen HomeScan sample of single households in 2010. The table presents the sample mean and standard deviation (in parenthesis).*

5.2.1. **Data description**

We use a subset of the data introduced in Harding and Lovenheim (2013), which draws on a large panel of household food purchases from the Nielsen Homescan database. The data records all purchases at the UPC level for a large sample of nationally representative households. Purchases are made in a variety of supermarkets and grocery stores and are meant for at home consumption. Each purchased product is uniquely identified through its Universal Product Code (UPC), a barcode, which is scanned at the point of sale. Nielsen requires all participating households to re-scan each purchased product at home. Each week the scanned data together with the receipts are transmitted to Nielsen where the data is combined and verified against store sale information. As a result the data contains accurate measures of the price and quantity of each purchased product.

For each household Nielsen collects detailed demographic information for the head of household using a yearly survey. The data includes the gender, race, income and education of the head of household.

While providing detailed information on each transaction, the Nielsen data does not record the nutritional content of every product. Harding and Lovenheim (2013) merge the transaction level data at the UPC level with detailed nutrition databases obtained from Gladson and FoodEssentials, which contain the exact nutrition panel of each product.

In this paper we restrict attention to single person households (Table 5.3). The sample includes 6326 female consumers and 2839 male consumers observed during 12 months in 2010. In the empirical analysis, we focus on six of the most essential (and salient) nutrients consumed per month: total fat, salt, sugar, cholesterol, protein, and carbohydrates.

5.2.2. *Empirical results*

In this analysis, we explore how preferences for nutrients vary across gender. Our quantile regression framework allows us to additionally investigate the extent to which gender differences are uniform over the conditional distribution of expenditure. In Table 5.4, we report the estimated coefficients for the QR, FEQR and SQR models for females. The profile of the asymptotic variance of the SQR suggests a value of $\lambda = 1.21$. We explore the results further in Figures 5.2 and 5.3. Each subfigure plots the estimated coefficient for a covariate of interest for each of the 0.1, 0.5, and 0.9 quantiles across a range of shrinkage parameters λ. In Figure 5.2 we report the estimated preferences for females, while Figure 5.3 reports the difference between the estimated coefficients for females and males.

First, consider the results for female shoppers. We find that higher food expenditures are associated with stronger preferences for fat, salt, sugar, and protein. These nutrients are highly correlated with taste and the richer a product is in these nutrients the more likely it is to be appealing to a consumer. In contrast, cholesterol and carbohydrates are generally associated with health hazards and have no direct impact on taste. As a result we see that high expenditure households prefer these nutrients less than low expenditure households, which may be indicative of avoidance behavior. Popular culture places a lot of emphasis on diets aimed at avoiding cholesterol and carbohydrates. It is also important to stress that the differences between the estimates for the extreme quantiles and the median reveal substantial preference heterogeneity across the conditional expenditure distribution which varies by nutrient. For instance, the effect of protein is negative and significant at the 0.1 quantile, while positive and significant at the 0.9 quantile. Furthermore, our approach allows us to

	Quantiles								
	0.10			0.50			0.90		
	QR	FEQR	SQR	QR	FEQR	SQR	QR	FEQR	SQR
Total fat	1.204	1.175	1.144	1.216	1.170	1.173	1.229	1.273	1.291
	(0.038)	(0.048)	(0.034)	(0.028)	(0.071)	(0.081)	(0.058)	(0.085)	(0.081)
Salt	0.003	0.002	0.002	0.003	0.003	0.003	0.005	0.004	0.004
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
Sugar	0.245	0.179	0.165	0.439	0.357	0.362	0.707	0.554	0.551
	(0.024)	(0.038)	(0.036)	(0.021)	(0.032)	(0.034)	(0.044)	(0.050)	(0.057)
Cholesterol	0.091	0.177	0.175	-0.045	0.061	0.064	-0.145	0.008	0.007
	(0.011)	(0.020)	(0.019)	(0.010)	(0.027)	(0.030)	(0.016)	(0.020)	(0.024)
Protein	-0.013	-0.120	-0.115	2.761	1.856	1.842	5.134	2.960	3.020
	(0.110)	(0.049)	(0.058)	(0.138)	(0.393)	(0.435)	(0.222)	(0.317)	(0.337)
Carbohydrates	0.414	0.552	0.542	0.356	0.437	0.438	0.331	0.388	0.406
	(0.017)	(0.025)	(0.022)	(0.019)	(0.039)	(0.044)	(0.033)	(0.042)	(0.038)
Black	-5.305		-6.135	-6.717		-4.700	-8.570		-3.893
	(0.427)		(1.058)	(0.410)		(1.493)	(0.884)		(1.922)
High income	6.368		7.134	13.990		12.915	22.152		20.305
	(0.561)		(1.581)	(0.539)		(1.865)	(0.974)		(2.198)
College education	3.218		3.979	6.421		5.761	10.707		8.468
	(0.319)		(1.089)	(0.316)		(1.350)	(0.653)		(1.691)

TABLE 5.4. *Estimating a panel quantile demand model for females. QR refers to quantile regression, FEQR is fixed effects quantile regression and SQR is the penalized estimator for the semiparametric correlated random effects model. The value of the tuning parameter is 1.21.*

estimate directly the impact of the demographics on food expenditure. We find that being African-American is associated with a negative impact on overall food expenditures, while households with incomes above \$70,000 and with a College education are associated with higher expenditures on food. These effects become more pronounced across the expenditure distribution. Note that the results of the SQR model are quantitatively very similar to those of the FEQR model, but substantially different than the QR results. This re-inforces the importance of controlling for unobserved heterogeneity. However, our approach also has the advantage of enabling us to derive the impact of the observable demographics on food expenditures.

Figure 5.3 reveals that these preferences also differ substantially between males and females. We find that for individuals at the lower tail of the conditional expenditure distribution, females have stronger preferences for fat than males, while the opposite is true for salt and
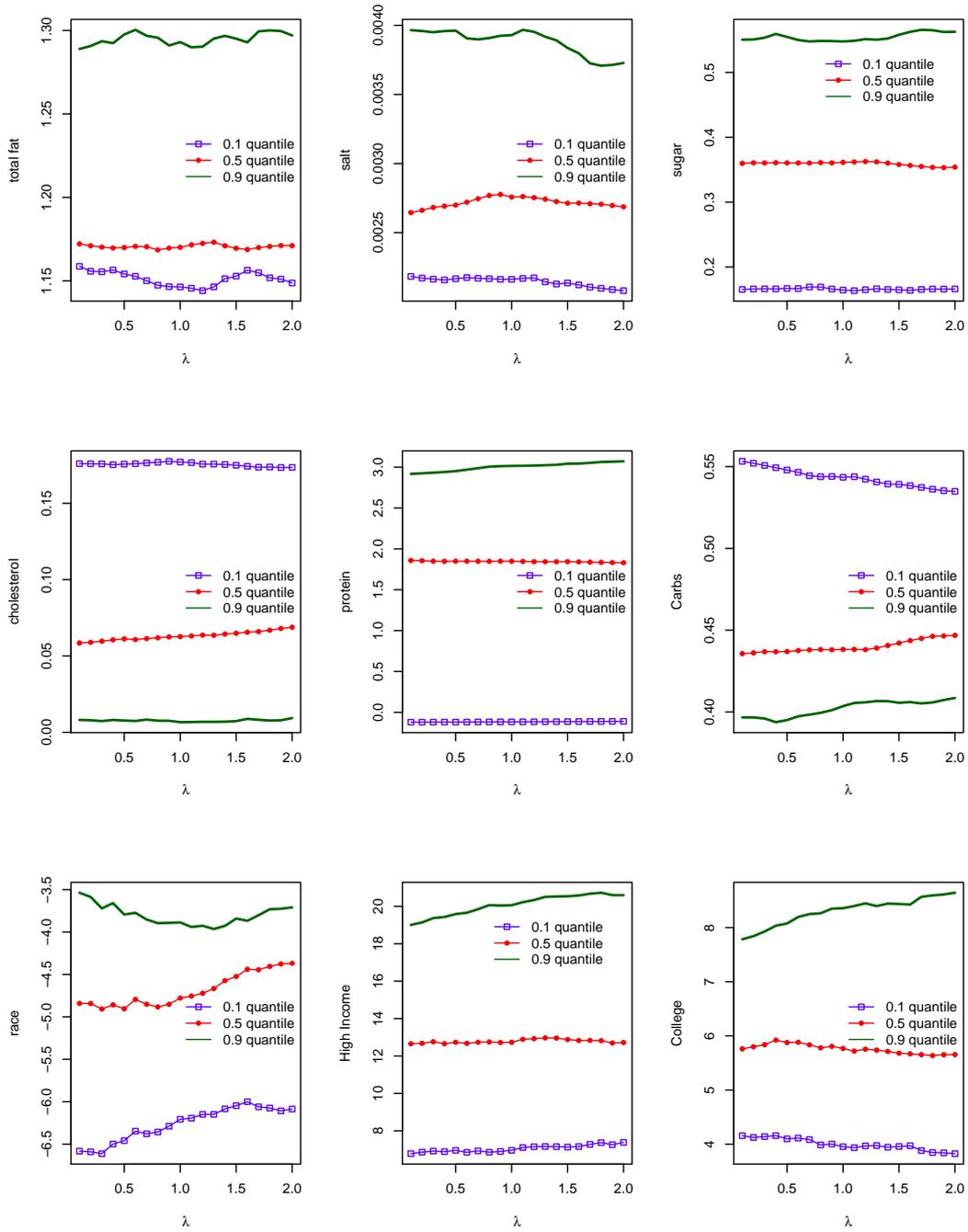
FIGURE 5.2. *Profile of preferences for female consumers.*
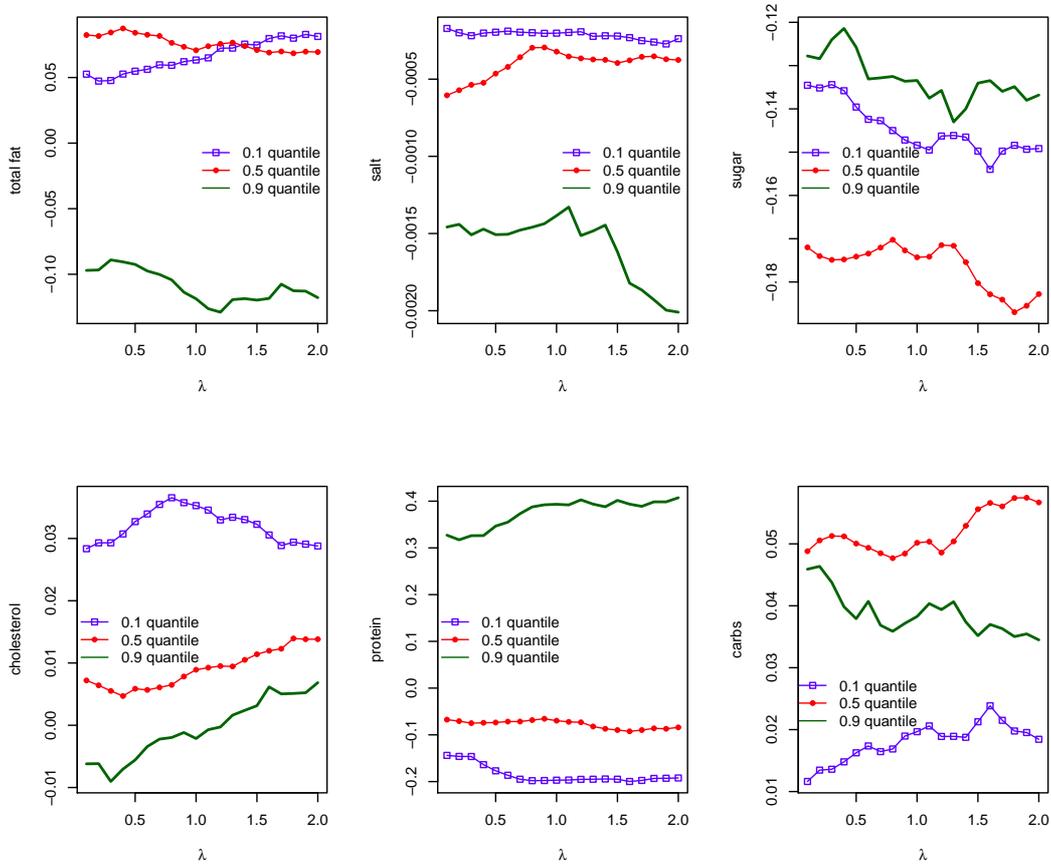
FIGURE 5.3. *Profile of preferences by gender in terms of λ.*

sugar. Furthermore, females have stronger preferences for protein at the upper quantiles of the distribution, while males prefer protein at the lower quantiles. The reverse appears to be true for cholesterol. This analysis reveals consumer preferences to be extremely heterogeneous and cautions against the use of demand models assuming homogeneous preferences.

From an econometric perspective it is also important to note the role played by the shrinkage parameter λ. In general we would expect the estimated coefficients on the covariates of interest to be fairly comparable across different degrees of shrinkage. By construction, the procedure shrinks the individual effects towards zero and if the model is well specified we would not expect changes in λ to affect the values of the estimated coefficients at each quantile. If on the other hand, the distribution of the individual effects is non-zero this indicates that the model retains additional unobserved variables with non-zero mean, which

are correlated with the covariates and are not captured by the flexible function $g(\cdot)$. Our estimates appear to indicate that while the model appears to fit the data well, there is the possibility of additional endogeneity not fully captured by the model specification as indicated by the estimates of the preferences for carbohydrates.

6. Conclusions

This paper investigates simple ℓ_1-penalized approaches to the estimation of marginal effects in panel data models with time-invariant variables by allowing for a flexible specification of correlated individual effects in a quantile regression setting. The approaches offer a balanced compromise between misspecification issues arising from the omission of individual heterogeneity and the incidental parameters problem arising from leaving individual heterogeneity unrestricted in a nonlinear panel model. We provide two empirical applications illustrating the practical implementation and use of the proposed methods.

In the first application, we study the responsiveness of hours to wages using BHPS data. The proposed method offers the possibility of estimating the main parameter of interest for policy analysis, the wage elasticity of labor supply, while allowing for heterogeneity in preferences. The findings show that: (i) positive substitution effects dominate negative wealth effects at several quantiles of the conditional distribution of hours; (ii) the overall positive effect of an increase in wages on hours tends to decrease across quantiles; (iii) wealth effects are small among workers having zero non-labor income, but they significantly affect workers with positive non-labor income.

In the second application we estimate consumer preferences for nutrients from a semi-structural demand model using a large scanner dataset of household food purchases. We show that preferences for nutrients vary across the conditional distribution of expenditure and across genders and emphasize the importance of fully capturing consumer heterogeneity in demand modeling and policy evaluation.

References

ABREVAYA, J., AND C. DAHL (2008): "The Effects of Smoking and Prenatal Care on Birth Outcomes: Evidence from Quantile Regression Estimation on Panel Data," *Journal of Business and Economics Statistics*, 26(4), 379–397.

BALTAGI, B. (2008): *Econometric Analysis of Panel Data*. Wiley, New York, 4 edn.

32

Battistin, E., and B. Sianesi (2011): "Misclassified Treatment Status and Treatment Effects: An Application to Returns to Education in the UK," *Review of Economic and Statistics*, 93 (2), 495–509.

Belloni, A., and V. Chernozhukov (2011): "L_1-Penalized Quantile Regression in High-Dimensional Sparse Models," *Annals of Statistics*, 39(1), 82–130.

Bhatia, R. (1997): *Matrix Analysis*. Springer-Verlag, New York.

Bickel, P., and B. Li (2006): "Regularization in Statistics," *Sociedad de Estadistica e Investigacion Operativa*, 15.

Blundell, R., and T. MaCurdy (1999): "Labor Supply: A Review of Alternative Approaches," in *Handbook of Labor Economics*, ed. by O. Ashenfelter, and D. Card, vol. 3, pp. 1560–1618. Elsevier Science.

Blundell, R., T. MaCurdy, and C. Meghir (2007): "Labor Supply Models: Unobserved Heterogeneity, Nonparticipation and Dynamics," in *Handbook of Econometrics*, ed. by J. Heckman, and E. Leamer, vol. 6A. Elsevier Science.

Blundell, R., and C. Meghir (1986): "Selection Criteria for A Microeconometric Model of Labour Supply," *Journal of Applied Econometrics*, 1, 55–80.

Burtless, G., and J. A. Hausman (1978): "The Effect of Taxation on Labor Supply: Evaluating the Gary Negative Income Tax Experiment," *Journal of Political Economy*, 86(6), pp. 1103–1130.

Cai, Z., and Z. Xiao (2012): "Semiparametric quantile regression estimation in dynamic models with partially varying coefficients," *Journal of Econometrics*, 167(2), 413 – 425.

Canay, I. A. (2011): "A simple approach to quantile regression for panel data," *The Econometrics Journal*, 14(3), 368–386.

Card, D. (1999): "The Causal Effect of Education on Earnings," in *Handbook of Labor Economics*, ed. by O. Ashenfelter, and D. Card, vol. 3. Elsevier Science.

Carrasco, M., J. P. Florens, and E. Renault (2007): "Linear Inverse Problems in Structural Econometrics: Estimation Based on Spectral Decomposition and Regularization," in *Handbook of Econometrics*, ed. by J. Heckman, and E. Leamer, vol. 6B. Elsevier Science.

Chamberlain, G. (1982): "Multivariate Regression Models for Panel Data," *Journal of Econometrics*, 18, 5–46.

Chen, X. (2007): "Large Sample Sieve Estimation of Semi-Nonparametric Models," *Handbook of Econometrics, Edited by J. Heckman and E. Leamer*, 6B, Elsevier.

Chernozhukov, V., I. Fernández-Val, J. Hahn, and W. Newey (2013): "Average and Quantile Effects in Nonseparable Panel Models," *Econometrica*, 81, 535–580.

Donoho, D., S. Chen, and M. Saunders (1998): "Atomic Decomposition by Basis Pursuit," *SIAM Journal of Scientific Computing*, 20, 33–61.

Dubois, P., R. Griffith, and A. Nevo (2013): "Do Prices and Attributes Explain International Differences in Food Purchases," *American Economic Review*, forthcoming.

Galvao, A., C. Lamarche, and L. Lima (2013): "Estimation of Censored Quantile Regression for Panel Data with Fixed Effects," Journal of the American Statistical Association, forthcoming.

Galvao, A. F. (2011): "Quantile regression for dynamic panel data with fixed effects," *Journal of Econometrics*, 164(1), 142 – 157.

GRAHAM, B. S., J. HAHN, AND J. L. POWELL (2009): "The Incidental Parameter Problem in a Non-differentiable Panel Data Model," *Economics Letters*, 105, 181–182.

HARDING, M., AND C. LAMARCHE (2009): "A Quantile Regression Approach for Estimating Panel Data Models Using Instrumental Variables," *Economics Letters*, 104, 133–135.

HARDING, M., AND C. LAMARCHE (2013): "Estimating and Testing a Quantile Regression Model with Interactive Effects," Journal of Econometrics, forthcoming.

HARDING, M., AND M. LOVENHEIM (2013): "The Effect of Product and Nutrient-Specific Taxes on Shopping Behavior and Nutrition: Evidence from Scanner Data," working paper, Stanford.

HAUSMAN, J. A. (1985): "Taxes and labor supply," vol. 1 of *Handbook of Public Economics*, chap. 4, pp. 213 – 263. Elsevier.

HE, X., AND P. SHI (1996): "Bivariate tensor-product B-splines in a partly linear model," *Journal of Multivariate Analysis*, 58, 162–181.

HOROWITZ, J., AND S. LEE (2007): "Nonparametric Instrumental Variables Estimation of a Quantile Regression Model," *Econometrica*, 75(4), 1191–1208.

HSIAO, C. (2003): *Analysis of Panel Data*. Cambridge University Press, New York, 2nd edn.

KATO, K., A. F. GALVAO, AND G. MONTES-ROJAS (2012): "Asymptotics for Panel Quantile Regression Models with Individual Effects," *Journal of Econometrics*, 170, 76–91.

KOENKER, R. (2004): "Quantile Regression for Longitudinal Data," *Journal of Multivariate Analysis*, 91, 74–89.

———— (2005): *Quantile Regression*. Cambridge University Press.

KOENKER, R., AND K. HALLOCK (2000): "Quantile Regression: An Introduction," mimeo, Department of Economics, University of Illinois at Urbana-Champaign.

KUMAR, A. (2012): "Nonparametric estimation of the impact of taxes on female labor supply," *Journal of Applied Econometrics*, 27(3), 415–439.

LAMARCHE, C. (2010): "Robust Penalized Quantile Regression Estimation for Panel Data," *Journal of Econometrics*, 157, 396–408.

MACURDY, T. (1981): "An Empirical Model of Labor Supply in a Life-Cycle Setting," *Journal of Political Economy*, 89, 1059–1085.

MUELLBAUER, J. (1974): "Household Production Theory, Quality, and the Hedonic Technique," *American Economic Review*, 64(6), 977–994.

NEYMAN, J., AND E. L. SCOTT (1948): "Consistent Estimates Based on Partially Consistent Observations," *Econometrica*, 16, 1–32.

OGDEN, C., M. CARROLL, B. KIT, AND K. FLEGAL (2012): "Prevalence of obesity in the United States," NCHS data brief, no 82.

PISTAFERRI, L. (2003): "Anticipated and Unanticipated Wage Changes, Wage Risk, and Intertemporal Labor Supply," *Journal of Labor Economics*, 21(3).

ROSEN, A. M. (2012): "Set identification via quantile restrictions in short panels," *Journal of Econometrics*, 166(1), 127 – 137.

TANG, Y., H. J. WANG, X. HE, AND Z. ZHU (2012): "An Informative Subset-Based Estimator for Censored Quantile Regression," *Test*, 21, 635–655.

TIBSHIRANI, R. (1996): "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society B*, 58, 267–288.

VAN SOEST, A., M. DAS, AND X. GONG (2002): "A Structural Labor Supply Model with Flexible Preferences," *Journal of Econometrics*, 107, 345–374.

WEI, Y., AND X. HE (2006): "Conditional Growth Charts," *Annals of Statistics*, 34(5), 2069–2097.

WEI, Y., A. PERE, R. KOENKER, AND X. HE (2006): "Quantile regression methods for reference growth charts," *Statistics in Medicine*, 25, 1369–1382.

ZILIAK, J. P., AND T. J. KNIESNER (1999): "Estimating Life Cycle Labor Supply Tax Effects," *Journal of Political Economy*, 107(2), pp. 326–359.

Appendix A. Technical Appendix

A.1. Some useful lemmas

LEMMA 1. *Let a $p \times p$ matrix $\boldsymbol{F} = (\boldsymbol{B} + \lambda \boldsymbol{I})^{-1} \boldsymbol{A}^{-1} \boldsymbol{C} (\boldsymbol{D} + \lambda^2 \boldsymbol{I})(\boldsymbol{B} + \lambda \boldsymbol{I})^{-1} \boldsymbol{A}^{-1}$, where $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, and $\boldsymbol{D}$ are $p \times p$ positive definite matrices with eigenvalues ζ_a, ζ_b, ζ_c, and ζ_d. Then, the trace of $\boldsymbol{F}$ is equal to,*

$$
tr\boldsymbol{F} = \sum_{i=1}^{p} \frac{\zeta_c^i (\zeta_d^i + \lambda^2)}{(\zeta_a^i (\zeta_b^i + \lambda))^2}.
$$

Proof. We consider a spectral decomposition for the matrices $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, and $\boldsymbol{D}$. The matrix $\boldsymbol{A} = \boldsymbol{U}_a \boldsymbol{\Lambda}_a \boldsymbol{U}_a'$, where $\boldsymbol{U}$ is an orthogonal matrix, and $\boldsymbol{\Lambda}$ is a diagonal matrix that contains the characteristic roots of matrix $\boldsymbol{A}$, with a typical element ζ_a^i for $i = 1, \ldots, p$. The trace of $\boldsymbol{F}$ is then,

$$
\begin{aligned}
tr\boldsymbol{F} &= tr\{(\boldsymbol{B} + \lambda \boldsymbol{I})^{-1} \boldsymbol{A}^{-1} \boldsymbol{C} (\boldsymbol{D} + \lambda^2 \boldsymbol{I})(\boldsymbol{B} + \lambda \boldsymbol{I})^{-1} \boldsymbol{A}^{-1}\} \\
&= tr\{(\boldsymbol{U}_b \boldsymbol{\Lambda}_b \boldsymbol{U}_b' + \lambda \boldsymbol{I})^{-1} (\boldsymbol{U}_a \boldsymbol{\Lambda}_a \boldsymbol{U}_a')^{-1} \boldsymbol{U}_c \boldsymbol{\Lambda}_c \boldsymbol{U}_c' (\boldsymbol{U}_d \boldsymbol{\Lambda}_d \boldsymbol{U}_d' + \lambda^2 \boldsymbol{I})(\boldsymbol{U}_b \boldsymbol{\Lambda}_b \boldsymbol{U}_b' + \lambda \boldsymbol{I})^{-1} (\boldsymbol{U}_a \boldsymbol{\Lambda}_a \boldsymbol{U}_a')^{-1}\} \\
&= tr\{\boldsymbol{U}_b' (\boldsymbol{\Lambda}_b + \lambda \boldsymbol{I})^{-1} \boldsymbol{U}_b \boldsymbol{U}_a' \boldsymbol{\Lambda}_a^{-1} \boldsymbol{U}_a \boldsymbol{U}_c' \boldsymbol{\Lambda}_c \boldsymbol{U}_c \boldsymbol{U}_d' (\boldsymbol{\Lambda}_d + \lambda^2 \boldsymbol{I}) \boldsymbol{U}_d \boldsymbol{U}_b' (\boldsymbol{\Lambda}_b + \lambda \boldsymbol{I})^{-1} \boldsymbol{U}_b \boldsymbol{U}_a' \boldsymbol{\Lambda}_a^{-1} \boldsymbol{U}_a\} \\
&= tr\left\{(\boldsymbol{\Lambda}_b + \lambda \boldsymbol{I})^{-1} \boldsymbol{\Lambda}_a^{-1} \boldsymbol{\Lambda}_c (\boldsymbol{\Lambda}_d + \lambda^2 \boldsymbol{I})(\boldsymbol{\Lambda}_b + \lambda \boldsymbol{I})^{-1} \boldsymbol{\Lambda}_a^{-1}\right\} \\
&= \sum_{i=1}^{p} \frac{\zeta_c^i (\zeta_d^i + \lambda^2)}{(\zeta_a^i (\zeta_b^i + \lambda))^2}.
\end{aligned}
$$

where the fourth equality holds because the $tr\boldsymbol{ABA} = tr\boldsymbol{AAB}$ and $\boldsymbol{U}'\boldsymbol{U} = \boldsymbol{I}$. $\qquad\square$

LEMMA 2. *Let $\boldsymbol{\Upsilon} = (\boldsymbol{I} - \boldsymbol{\Lambda})\boldsymbol{\Phi}$ and $\boldsymbol{P}_{\boldsymbol{\Upsilon}} = (\boldsymbol{Z}'\boldsymbol{\Upsilon}\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{\Upsilon}$. Then the weighted projection matrices $\boldsymbol{P}_{\boldsymbol{\Phi}}$ and $\boldsymbol{P}_{\boldsymbol{\Upsilon}}$ achieve the same transformation.*

Proof. Write $\boldsymbol{P_\Upsilon X} = (\boldsymbol{Z'\Upsilon Z})^{-1}\boldsymbol{Z'\Upsilon X}$. Dividing and multiplying $\boldsymbol{P_\Upsilon X}$ by T and using the fact that $T\boldsymbol{I} = \boldsymbol{ZZ'}$,

$$
\begin{aligned}
\boldsymbol{P_\Upsilon X} &= (\boldsymbol{Z'(I - \Lambda)ZZ'\Phi Z})^{-1}\boldsymbol{Z'(I - \Lambda)ZZ'\Phi X} \\
&= (\boldsymbol{Z'\Phi Z})^{-1}(\boldsymbol{Z'(I - \Lambda)Z})^{-1}\boldsymbol{Z'(I - \Lambda)ZZ'\Phi X} = (\boldsymbol{Z'\Phi Z})^{-1}\boldsymbol{Z'\Phi X} = \boldsymbol{P_\Phi X}.
\end{aligned}
$$

$\square$

A.2. Assumptions and sketch of the proofs

This paper employs the following regularity conditions:

ASSUMPTION 1. *The variables y_{it} are independent with conditional distribution $F_{Y_{it}}$, and continuous densities f_{it} uniformly bounded away from 0 and ∞, with bounded derivatives f'_{it}, at the points $\xi_{it}(\tau_j)$ for $j = 1, \ldots, J$, $t = 1, \ldots, T$ and $i = 1, \ldots, N$.*

ASSUMPTION 2. *The random variables a_i are identically, and independently distributed with unconditional distribution function F_a with median zero, and continuous densities f_a uniformly bounded away from 0 and ∞, with bounded derivatives f'_a, for $i = 1, \ldots, N$.*

ASSUMPTION 3. *There exist positive definite matrices $\boldsymbol{\Sigma}_0$, $\boldsymbol{\Sigma}_1$, $\boldsymbol{\Sigma}_2$, and $\boldsymbol{\Sigma}_3$ such that*

$$
\boldsymbol{\Sigma}_0 = \lim_{\substack{T\to\infty \\ N\to\infty}} \frac{1}{TN}
\begin{bmatrix}
\Omega_{11}\boldsymbol{X'W'_1W_1X} & \cdots & \Omega_{1J}\boldsymbol{X'W'_1W_JX} \\
\vdots & \ddots & \vdots \\
\Omega_{1J}\boldsymbol{X'W'_JW_1X} & \cdots & \Omega_{JJ}\boldsymbol{X'W'_JW_JX}
\end{bmatrix}
$$

$$
\boldsymbol{\Sigma}_1 = \lim_{\substack{T\to\infty \\ N\to\infty}} \frac{1}{TN}
\begin{bmatrix}
\omega_1\boldsymbol{X'W'_1\Upsilon_1W_1X} & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & \omega_J\boldsymbol{X'W'_J\Upsilon_JW_JX}
\end{bmatrix}
$$

$$
\boldsymbol{\Sigma}_2 = \lim_{\substack{T\to\infty \\ N\to\infty}} \frac{\Omega_m}{TN}
\begin{bmatrix}
\boldsymbol{X'P'_1P_1X} & \cdots & \boldsymbol{X'P'_1P_JX} \\
\vdots & \ddots & \vdots \\
\boldsymbol{X'P'_JP_1X} & \cdots & \boldsymbol{X'P'_JP_JX}
\end{bmatrix}
$$

$$
\boldsymbol{\Sigma}_3 = \lim_{\substack{T\to\infty \\ N\to\infty}} \frac{1}{NT}
\begin{bmatrix}
\boldsymbol{X'P'_1\Psi P_1X} & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & \boldsymbol{X'P'_J\Psi P_JX}
\end{bmatrix}
$$

where $\Omega_{kl} = \omega_k(\tau_k \wedge \tau_l - \tau_k\tau_l)\omega_l$ and $\Omega_m = \tau_m(1 - \tau_m)$ for the median τ_m; $\boldsymbol{W}_j = \boldsymbol{I} - \boldsymbol{ZP}_j$, $\boldsymbol{P}_j = (\boldsymbol{Z'\Upsilon_jZ})^{-1}\boldsymbol{Z'\Upsilon_j}$, $\boldsymbol{\Upsilon}_j = \boldsymbol{\Phi}_j(\boldsymbol{I} - \boldsymbol{\Lambda}_j)$, $\boldsymbol{\Phi}_j = diag(f_{it}(\xi_{it}(\tau_j)))$, $\boldsymbol{\Lambda}_j = diag(\boldsymbol{x}'_i(\boldsymbol{D'Z'\Phi_jZD})^{-1}\boldsymbol{\Phi}_{ij}\boldsymbol{x}_i)$,

$\boldsymbol{\Psi} = diag(f_a(0)),$ *and*

$$
\boldsymbol{X} = \begin{bmatrix} \boldsymbol{x}'_{11} \\ \boldsymbol{x}'_{12} \\ \vdots \\ \boldsymbol{x}'_{NT} \end{bmatrix} ; \boldsymbol{D} = \begin{bmatrix} \boldsymbol{x}'_{11} & \boldsymbol{x}'_{12} & ... & \boldsymbol{x}'_{1T} \\ \boldsymbol{x}'_{21} & \boldsymbol{x}'_{22} & ... & \boldsymbol{x}'_{2T} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{x}'_{N1} & \boldsymbol{x}'_{N2} & ... & \boldsymbol{x}'_{NT} \end{bmatrix} ; \boldsymbol{Z} = \begin{bmatrix} 1 & 0 & ... & 0 \\ 1 & 0 & ... & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & ... & 1 \end{bmatrix}.
$$

ASSUMPTION 4. $max||\boldsymbol{x}_{it}||/\sqrt{TN} \to 0$.

ASSUMPTION 5. *There exists a constant $c > 0$ such that $N^c/T \to 0$.*

ASSUMPTION 6. *The shrinkage parameter $\lambda_T/\sqrt{T} \to \lambda \geq 0$.*

Proof of Theorem 1. Let

$$
V_{NT}(\boldsymbol{\delta}) = \sum_{j=1}^{J}\sum_{t=1}^{T}\sum_{i=1}^{N} \omega_j \rho_{\tau_j}(y_{it} - \xi_{it}(\tau_j) - \delta_{0i}/\sqrt{T} - \boldsymbol{x}'_{it}\boldsymbol{\delta}_1(\tau_j)/\sqrt{NT} - \boldsymbol{x}'_i\boldsymbol{\delta}_2(\tau_j)/\sqrt{N})
$$

$$
- \rho_{\tau_j}(y_{it} - \xi_{it}(\tau_j)) + \lambda_T \sum_{i=1}^{N}\rho_{\tau_m}(a_i + \delta_{0i}/\sqrt{T}) - \rho_{\tau_m}(a_i)
$$

where τ_m is the median quantile and $\xi_{it}(\tau_j) = \boldsymbol{x}'_{it}\boldsymbol{\beta}(\tau_j)+\boldsymbol{x}'_i\boldsymbol{\gamma}(\tau_j)+a_i$ is the conditional quantile function. We overcome the difficulty associated with infinite dimensional vectors by concentrating out the δ_{2t}'s and δ_{0i}'s effects into the objective function. For any $(\Delta_{0i}, \Delta_1, \Delta_2) > 0$, $\sup ||k(\delta_{0i}, \boldsymbol{\delta}_1, \boldsymbol{\delta}_2) - k(0, \boldsymbol{0}, \boldsymbol{0}) - \mathbb{E}(k(\delta_{0i}, \boldsymbol{\delta}_1, \boldsymbol{\delta}_2) - k(0, \boldsymbol{0}, \boldsymbol{0}))|| = o_p(1)$, where,

$$
k_t(\delta_{0i}, \boldsymbol{\delta}_1, \boldsymbol{\delta}_2) = -\frac{1}{\sqrt{N}}\sum_{j=1}^{J}\sum_{i=1}^{N}\omega_j\boldsymbol{x}_i\psi_{\tau_j}\left(y_{it} - \frac{\delta_{0i}}{\sqrt{T}} - \boldsymbol{x}'_{it}\frac{\boldsymbol{\delta}_1(\tau_j)}{\sqrt{TN}} - \boldsymbol{x}'_i\frac{\boldsymbol{\delta}_2(\tau_j)}{\sqrt{N}} - \xi_{it}(\tau_j)\right)
$$

with $\psi_{\tau_j}(u) = \tau_j - I(u < 0)$. Taking expectation and expanding $k(\cdot)$ under condition 1, we obtain

$\mathbb{E}((k_t(\delta_{0i}, \boldsymbol{\delta}_1, \boldsymbol{\delta}_2)) - k_t(0, \boldsymbol{0}, \boldsymbol{0}))) =$

$$
= \mathbb{E}\left(-\frac{1}{\sqrt{N}}\sum_{j=1}^{J}\sum_{i=1}^{N}\omega_j\boldsymbol{x}_i\left(\psi_{\tau_j}\left(y_{it} - \frac{\delta_{0i}}{\sqrt{T}} - \boldsymbol{x}'_{it}\frac{\boldsymbol{\delta}_1(\tau_j)}{\sqrt{TN}} - \boldsymbol{x}'_i\frac{\boldsymbol{\delta}_2(\tau_j)}{\sqrt{N}} - \xi_{it}(\tau_j)\right) - \psi_{\tau_j}(y_{it} - \xi_{it}(\tau_j))\right)\right)
$$

$$
= -\frac{1}{\sqrt{N}}\sum_{j=1}^{J}\sum_{i=1}^{N}\omega_j\boldsymbol{x}_i\left(F_{it}\left(\xi_{it}(\tau_j) + \frac{\delta_{0i}}{\sqrt{T}} + \boldsymbol{x}'_{it}\frac{\boldsymbol{\delta}_1(\tau_j)}{\sqrt{TN}} + \boldsymbol{x}'_i\frac{\boldsymbol{\delta}_2(\tau_j)}{\sqrt{N}}\right) - \tau_j\right)
$$

$$
= -\frac{1}{\sqrt{N}}\sum_{j=1}^{J}\sum_{i=1}^{N}\omega_j f_{it}(\xi_{it}(\tau_j))\boldsymbol{x}_i\left(\frac{\delta_{0i}}{\sqrt{T}} + \boldsymbol{x}'_{it}\frac{\boldsymbol{\delta}_1(\tau_j)}{\sqrt{TN}} + \boldsymbol{x}'_i\frac{\boldsymbol{\delta}_2(\tau_j)}{\sqrt{N}}\right) + o_p(1)
$$

Optimality of the $\hat{\delta}_{2t}$'s implies that $k_t(\delta_{0i}, \boldsymbol{\delta}_1, \boldsymbol{\delta}_2) = o(N^{-1})$, and thus $\mathbb{E}(k_t(\delta_{0i}, \boldsymbol{\delta}_1, \boldsymbol{\delta}_2) - k_t(0, \boldsymbol{0}, \boldsymbol{0})) = k_t(0, \boldsymbol{0}, \boldsymbol{0})$. This last expression can be written as,

$$\frac{1}{\sqrt{N}} \sum_{j=1}^{J} \sum_{i=1}^{N} \omega_j \boldsymbol{x}_i f_{it}(\xi_{it}(\tau_j)) \left(\frac{\delta_{0i}}{\sqrt{T}} + \boldsymbol{x}_{it}' \frac{\boldsymbol{\delta}_1(\tau_j)}{\sqrt{TN}} + \boldsymbol{x}_i' \frac{\boldsymbol{\delta}_2(\tau_j)}{\sqrt{N}} \right) = \frac{1}{\sqrt{N}} \sum_{j=1}^{J} \sum_{i=1}^{N} \omega_j \boldsymbol{x}_i \psi_{\tau_j} \left(y_{it} - \xi_{it}(\tau_j) \right)$$

Solving for $\boldsymbol{\delta}_2$,

$$\frac{\hat{\boldsymbol{\delta}}_2(\tau_j)}{\sqrt{N}} = -\boldsymbol{h}_{jit}^{-1} \left[\sum_{j=1}^{J} \sum_{i=1}^{N} \omega_j \check{\boldsymbol{x}}_{ij} \left(\frac{\delta_{0i}}{\sqrt{T}} + \boldsymbol{x}_{it}' \frac{\boldsymbol{\delta}_1(\tau_j)}{\sqrt{TN}} \right) + \sum_{j=1}^{J} \sum_{i=1}^{N} \omega_j \boldsymbol{x}_i \psi_{\tau_j}(y_{it} - \xi_{it}(\tau_j)) \right] + \frac{\boldsymbol{R}_{Nt}}{\sqrt{N}}$$

where $\boldsymbol{h}_{jit} = \sum_j \sum_i \omega_j \boldsymbol{x}_i \check{\boldsymbol{x}}_{ij}'$, $\check{\boldsymbol{x}}_{ij} = (f_{it}(\xi_{it}(\tau_j)) \boldsymbol{x}_{it})_{itj}$ and $\boldsymbol{R}$ is the remainder term. Substituting the $\hat{\delta}_{2t}$'s, we denote

$$k(\delta_{0i}, \boldsymbol{\delta}_1) = -\frac{1}{\sqrt{T}} \sum_{j=1}^{J} \sum_{t=1}^{T} \omega_j \psi_{\tau_j} \left(y_{it} - \frac{\delta_{0i}}{\sqrt{T}} - \boldsymbol{x}_{it}' \frac{\boldsymbol{\delta}_1(\tau_j)}{\sqrt{TN}} - \boldsymbol{x}_i' \frac{\hat{\boldsymbol{\delta}}_2(\tau_j)}{\sqrt{N}} - \xi_{it}(\tau_j) \right) + 2 \frac{\lambda_T}{\sqrt{T}} \psi_{\tau_m} \left(a_i + \frac{\delta_{0i}}{\sqrt{T}} \right)$$

Taking expectation and expanding as above, we obtain under conditions 1-2,

$$\mathbb{E}((k(\delta_{0i}, \boldsymbol{\delta}_1)) - k(0, \boldsymbol{0}))) =$$

$$= \mathbb{E} \Bigg(-\frac{1}{\sqrt{T}} \sum_{j=1}^{J} \sum_{t=1}^{T} \omega_j (\psi_{\tau_j}(y_{it} - \frac{\delta_{0i}}{\sqrt{T}} - \boldsymbol{x}_{it}' \frac{\boldsymbol{\delta}_1(\tau_j)}{\sqrt{TN}} - \boldsymbol{x}_i' \frac{\hat{\boldsymbol{\delta}}_2(\tau_j)}{\sqrt{N}} - \xi_{it}(\tau_j)) + 2 \frac{\lambda_T}{\sqrt{T}} \psi_{\tau_m} \left(a_i + \frac{\delta_{0i}}{\sqrt{T}} \right)$$

$$+ \frac{1}{\sqrt{T}} \sum_{j=1}^{J} \sum_{t=1}^{T} \omega_j \psi_{\tau_j}(y_{it} - \xi_{it}(\tau_j)) - 2 \frac{\lambda_T}{\sqrt{T}} \psi_{\tau_m}(a_i) \Bigg)$$

$$= -\frac{1}{\sqrt{T}} \sum_{j=1}^{J} \sum_{t=1}^{T} \omega_j \left(F_{it} \left(\xi_{it}(\tau_j) + \frac{\delta_{0i}}{\sqrt{T}} + \boldsymbol{x}_{it}' \frac{\boldsymbol{\delta}_1(\tau_j)}{\sqrt{TN}} + \boldsymbol{x}_i' \frac{\hat{\boldsymbol{\delta}}_2(\tau_j)}{\sqrt{N}} \right) - \tau_j \right) + 2 \frac{\lambda_T}{\sqrt{T}} \left(F_a \left(-\frac{\delta_{0i}}{\sqrt{T}} \right) - \frac{1}{2} \right)$$

$$= -\frac{1}{\sqrt{T}} \sum_{j=1}^{J} \sum_{t=1}^{T} \omega_j f_{it}(\xi_{it}(\tau_j)) \left(\frac{\delta_{0i}}{\sqrt{T}} + \boldsymbol{x}_{it}' \frac{\boldsymbol{\delta}_1(\tau_j)}{\sqrt{TN}} + \boldsymbol{x}_i' \frac{\hat{\boldsymbol{\delta}}_2(\tau_j)}{\sqrt{N}} \right) - 2 \frac{\lambda_T}{\sqrt{T}} f_a(0) \frac{\delta_{0i}}{\sqrt{T}} + o_p(1)$$

$$= -\frac{1}{\sqrt{T}} \sum_{j=1}^{J} \sum_{t=1}^{T} \omega_j f_{it}(\xi_{it}(\tau_j)) \left(\frac{\delta_{0i}}{\sqrt{T}} + \boldsymbol{x}_{it}' \frac{\boldsymbol{\delta}_1(\tau_j)}{\sqrt{TN}} - \boldsymbol{x}_i' \boldsymbol{h}_{jit}^{-1} \sum_{j=1}^{J} \sum_{i=1}^{N} \omega_j \check{\boldsymbol{x}}_{ij} \left(\frac{\delta_{0i}}{\sqrt{T}} + \boldsymbol{x}_{it}' \frac{\boldsymbol{\delta}_1(\tau_j)}{\sqrt{TN}} \right) \right.$$

$$\left. - \boldsymbol{x}_i' \boldsymbol{h}_{jit}^{-1} \sum_{j=1}^{J} \sum_{i=1}^{N} \omega_j \boldsymbol{x}_i \psi_{\tau_j}(y_{it} - \xi_{it}(\tau_j)) + \boldsymbol{x}_i' \frac{\boldsymbol{R}_{Nt}}{\sqrt{N}} \right) - 2 \frac{\lambda_T}{\sqrt{T}} f_a(0) \frac{\delta_{0i}}{\sqrt{T}} + o_p(1)$$

$$= -\frac{1}{\sqrt{T}} \sum_{j=1}^{J} \sum_{t=1}^{T} \omega_j f_{it}(\xi_{it}(\tau_j)) \left(\left(w_{itj} \frac{\delta_{0i}}{\sqrt{T}} + \mu_{jit} \boldsymbol{x}_{it}' \frac{\boldsymbol{\delta}_1(\tau_j)}{\sqrt{TN}} \right) - \boldsymbol{x}_i' \boldsymbol{h}_{jit}^{-1} \omega_j \boldsymbol{x}_i \psi_{\tau_j}(y_{it} - \xi_{it}(\tau_j)) + \boldsymbol{x}_i' \frac{\boldsymbol{R}_{Nt}}{\sqrt{N}} \right)$$

where $\mu_{jit} = 1 - \boldsymbol{x}_i' \boldsymbol{h}_{jit}^{-1} \check{\boldsymbol{x}}_{ij}$, $f_i = T^{-1} \sum_{j=1}^{J} \sum_{t=1}^{T} \omega_j f_{it}(\xi_{it}(\tau_j))$, and $w_{itj} = \mu_{itj} + f_i^{-1} \lambda_T / \sqrt{T} f_a(0)$. Optimality of the $\hat{\delta}_{0i}$'s implies that $k(\delta_{0i}, \boldsymbol{\delta}_1) = o(T^{-1})$, and thus $\mathbb{E}(k(\delta_{0i}, \boldsymbol{\delta}_1) - k(0, \boldsymbol{0})) = k(0, \boldsymbol{0})$. Therefore,

$$\frac{1}{\sqrt{T}} \sum_{j=1}^{J} \sum_{t=1}^{T} \omega_j f_{it} w_{itj} \frac{\delta_{0i}}{\sqrt{T}} + \frac{1}{\sqrt{T}} \sum_{j=1}^{J} \sum_{t=1}^{T} \omega_j f_{it} \mu_{jit} \boldsymbol{x}_{it}' \frac{\boldsymbol{\delta}_1(\tau_j)}{\sqrt{TN}}$$

$$= \frac{1}{\sqrt{T}} \sum_{j=1}^{J} \sum_{t=1}^{T} \omega_j f_{it} \mu_{itj} \psi_{\tau_j}(y_{it} - \xi_{it}(\tau_j)) + 2 \frac{\lambda_T}{\sqrt{T}} \psi_{\tau_m}(a_i) - \frac{1}{\sqrt{T}} \sum_{j=1}^{J} \sum_{t=1}^{T} \omega_j f_{it} \boldsymbol{x}_i' \frac{\boldsymbol{R}_{Nt}}{\sqrt{N}} + \frac{\boldsymbol{R}_{Ti}}{\sqrt{T}}$$

The asymptotic (Bahadur) representation of the individual specific effect relates to the slope parameter in the following way,

$$\frac{\hat{\delta}_{0i}}{\sqrt{T}} = -f_{ji}^{-1} \left[\frac{1}{T} \sum_{j=1}^{J} \sum_{t=1}^{T} \omega_j f_{it}(\xi_{it}(\tau_j)) \mu_{itj} \boldsymbol{x}_{it}' \right] \frac{\boldsymbol{\delta}_1(\tau_j)}{\sqrt{TN}} + f_{ji}^{-1} \left(\frac{1}{T} \sum_{j=1}^{J} \sum_{t=1}^{T} \omega_j f_{it} \mu_{itj} \psi_{\tau_j}(y_{it} - \xi_{it}(\tau_j)) \right.$$

$$\left. + 2 \frac{\lambda_T}{\sqrt{T}} \frac{\psi_{\tau_m}(a_i)}{\sqrt{T}} - \frac{1}{T} \sum_{j=1}^{J} \sum_{t=1}^{T} \omega_j f_{it} \boldsymbol{x}_i' \frac{\boldsymbol{R}_{Nt}}{\sqrt{N}} \right) + \frac{\boldsymbol{R}_{Ti}}{\sqrt{T}} = -\sum_{j=1}^{J} \tilde{\boldsymbol{x}}_{it}(\tau_j)' \frac{\boldsymbol{\delta}_1(\tau_j)}{\sqrt{TN}} + \boldsymbol{r}_{it},$$

where $f_{ji} = T^{-1} \sum_j \sum_t \omega_j f_{it} w_{itj}$ and $\tilde{\boldsymbol{x}}_{it} = \sum_t \omega_j f_{it} \mu_{itj} \boldsymbol{x}_{it}' / T f_i$. The term $\boldsymbol{r}_{it}$ includes the last four terms of the right hand side of the Bahadur presentation of the individual effects. By Lemma 1 in Lamarche (2010), the terms involving $\boldsymbol{r}_{it}$ converge to zero when $\hat{\delta}_{0i}/\sqrt{T}$ is inserted in the objective function. This requires T growing faster than N, which is satisfied for some values of c in Assumption 5 (See Kato, Galvao and Montes-Rojas (2012) for rates of convergence under fairly general conditions).

We now replace the asymptotic representation of the individual specific effect in the objective function, and decompose the equation in four terms defined as,

$$V_{TN}^{(1)}(\boldsymbol{\delta}_1) = -\sum_{j=1}^{J} \sum_{t=1}^{T} \sum_{i=1}^{N} \omega_j (\boldsymbol{x}_{it}' - \tilde{\boldsymbol{x}}_i(\tau_j)')(\boldsymbol{\delta}_1(\tau_j)/\sqrt{NT}) \psi_{\tau_j}(y_{it} - \xi_{it}(\tau_j))$$

$$V_{TN}^{(2)}(\boldsymbol{\delta}_1) = \sum_{j=1}^{J} \sum_{t=1}^{T} \sum_{i=1}^{N} \omega_j \int_0^{v_{itj,TN}} (I(y_{it} - \xi_{it}(\tau_j) \le s) - I(y_{it} - \xi_{it}(\tau_j) \le 0)) ds$$

$$V_{TN}^{(3)}(\boldsymbol{\delta}_1) = -\lambda_T \sum_{j=1}^{J} \sum_{i=1}^{N} \tilde{\boldsymbol{x}}_i(\tau_j)'(\boldsymbol{\delta}_1(\tau_j)/\sqrt{NT}) \psi_{\tau_m}(a_i)$$

$$V_{TN}^{(4)}(\boldsymbol{\delta}_1) = \lambda_T \sum_{i=1}^{N} \int_0^{\tilde{\boldsymbol{x}}_i' \boldsymbol{\delta}_1/\sqrt{TN}} (I(a_i \le s) - I(a_i \le 0)) ds$$

with $v_{itj,TN} = (\boldsymbol{x}'_{it} - \tilde{\boldsymbol{x}}_i(\tau_j)')\boldsymbol{\delta}_1(\tau_j)/\sqrt{TN}$. The first term is asymptotically Gaussian. By the Lindeberg-Feller Central Limit Theorem, and conditions 3-4,

$$V_{TN}^{(1)}(\boldsymbol{\delta}_1) = -\frac{1}{\sqrt{TN}} \sum_{j=1}^{J} \sum_{t=1}^{T} \sum_{i=1}^{N} \omega_j (\boldsymbol{x}'_{it} - \tilde{\boldsymbol{x}}_i(\tau_j)')\boldsymbol{\delta}_1(\tau_j)\psi_{\tau_j}(y_{it} - \xi_{it}(\tau_j)) \rightsquigarrow -\boldsymbol{\delta}'_1\mathbb{B}$$

The second term converges in probability to a quadratic term in $\boldsymbol{\delta}_1$,

$$\mathbb{E}V_{TN}^{(2)}(\boldsymbol{\delta}_1) = \frac{1}{2TN} \sum_{j=1}^{J} \sum_{t=1}^{T} \sum_{i=1}^{N} \omega_j f_{it}(\xi_{it}(\tau_j))((\boldsymbol{x}'_{it} - \tilde{\boldsymbol{x}}_i(\tau_j)')\boldsymbol{\delta}_1(\tau_j))^2 + o(1) \to \frac{1}{2}\boldsymbol{\delta}'_1\boldsymbol{\Sigma}_1\boldsymbol{\delta}_1$$

The variance of $V_{TN}^{(2)}(\boldsymbol{\delta}_1)$ converges to zero by condition 4. Similarly, by the Lindeberg-Feller Central Limit Theorem, the Slutsky Theorem, and conditions 3-4, the third term is asymptotically Gaussian,

$$V_{TN}^{(3)}(\boldsymbol{\delta}_1) = -\frac{\lambda_T}{\sqrt{T}} \frac{1}{\sqrt{N}} \sum_{j=1}^{J} \sum_{i=1}^{N} \tilde{\boldsymbol{x}}_i(\tau_j)'\boldsymbol{\delta}_1(\tau_j)\psi_{\tau_m}(a_i) \rightsquigarrow -\lambda\boldsymbol{\delta}'_1\mathbb{C},$$

where $\mathbb{C}$ is a Gaussian vector independent of $\mathbb{B}$ with covariance $\boldsymbol{\Sigma}_2$. The last last term has a quadratic contribution,

$$\mathbb{E}\left(V_{TN}^{(4)}(\boldsymbol{\delta}_1)\right) = \frac{\lambda_T}{2TN} \sum_{i=1}^{N} f_a(0)(\tilde{\boldsymbol{x}}'_i\boldsymbol{\delta}_1)^2 + o(1) \to \frac{1}{2}\lambda\boldsymbol{\delta}'_1\boldsymbol{\Sigma}_3\boldsymbol{\delta}_1$$

It follows that $V_{TN}(\boldsymbol{\delta}_1)$ is convex and $V_0(\boldsymbol{\delta}_1)$ has a unique minimum, and then $\text{argmin}(V_{TN}(\boldsymbol{\delta}_1(\tau))) \rightsquigarrow \text{argmin}(V_0(\boldsymbol{\delta}_1(\tau)))$. Therefore,

$$\sqrt{NT}(\hat{\boldsymbol{\beta}}(\tau,\lambda) - \boldsymbol{\beta}(\tau)) \rightsquigarrow (\boldsymbol{\Sigma}_1 + \lambda\boldsymbol{\Sigma}_3)^{-1}(\mathbb{B} + \lambda\mathbb{C}).$$

The penalized estimator converges to a Gaussian random variable with mean zero and covariance $\text{Avar}(\sqrt{NT}(\hat{\boldsymbol{\beta}}(\tau,\lambda)) = (\boldsymbol{\Sigma}_1 + \lambda\boldsymbol{\Sigma}_3)^{-1}(\boldsymbol{\Sigma}_0 + \lambda^2\boldsymbol{\Sigma}_2)(\boldsymbol{\Sigma}_1 + \lambda\boldsymbol{\Sigma}_3)^{-1}$.

We now derive the distribution of $\tilde{\boldsymbol{\beta}}(\tau)$. Consider the following objective function for the penalized estimator that shrinks endogenous α_i's as,

$$V_{NT}(\boldsymbol{\eta}) = \sum_{t=1}^{T} \sum_{i=1}^{N} \rho_\tau\left(y_{it} - \kappa_{it}(\tau) - \frac{\eta_{0i}}{\sqrt{T}} - \boldsymbol{x}'_{it}\frac{\boldsymbol{\eta}_1(\tau)}{\sqrt{NT}}\right) - \rho_\tau(y_{it} - \kappa_{it}(\tau)) + \lambda_T \sum_{i=1}^{N} \rho_\tau\left(\alpha_i + \frac{\eta_{0i}}{\sqrt{T}}\right) - \rho_\tau(\alpha_i)$$

where τ is the median quantile and $\kappa_{it}(\tau) = \boldsymbol{x}'_{it}\boldsymbol{\beta}(\tau) + \alpha_i$ is the conditional quantile function. Without loss of generality, we consider the location $s_i = \boldsymbol{x}'_i\boldsymbol{\gamma}$ in a neighborhood of 0. For any $(\Delta_{0i}, \Delta_1) > 0$,

$$\sup_{|\eta_{0i}|<\Delta_0, \|\eta_1\|<\Delta_1} ||v(\eta_{0i}, \boldsymbol{\eta}_1) - v(0, \boldsymbol{0}) - E(v(\eta_{0i}, \boldsymbol{\eta}_1) - v(0, \boldsymbol{0}))|| = o_p(1)$$

where,

$$v(\eta_{0i}, \boldsymbol{\eta}_1) = -\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \psi_\tau \left(y_{it} - \frac{\eta_{0i}}{\sqrt{T}} - \boldsymbol{x}_{it}' \frac{\boldsymbol{\eta}_1(\tau)}{\sqrt{TN}} - \kappa_{it}(\tau) \right) + 2\frac{\lambda_T}{\sqrt{T}} \psi_\tau \left(\alpha_i + \frac{\eta_{0i}}{\sqrt{T}} \right)$$

Taking expectation and expanding the function v as in Theorem 1, we obtain

$$\mathbb{E}((v(\eta_{0i}, \boldsymbol{\eta}_1)) - v(0, \mathbf{0}))) =$$

$$= \mathbb{E}\Bigg(-\frac{1}{\sqrt{T}} \sum_{t=1}^{T} (\psi_\tau \left(y_{it} - \frac{\eta_{0i}}{\sqrt{T}} - \boldsymbol{x}_{it}' \frac{\boldsymbol{\eta}_1(\tau)}{\sqrt{TN}} - \kappa_{it}(\tau) \right) + 2\frac{\lambda_T}{\sqrt{T}} \psi_\tau \left(\alpha_i + \frac{\eta_{0i}}{\sqrt{T}} \right)$$

$$+ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \psi_\tau \left(y_{it} - \kappa_{it}(\tau) \right) - 2\frac{\lambda_T}{\sqrt{T}} \psi_\tau(\alpha_i) \Bigg)$$

$$= -\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(F_{it} \left(\kappa_{it}(\tau) + \frac{\eta_{0i}}{\sqrt{T}} + \boldsymbol{x}_{it}' \frac{\boldsymbol{\eta}_1(\tau)}{\sqrt{TN}} \right) - \tau \right) + 2\frac{\lambda_T}{\sqrt{T}} \left(\tau - F_a \left(-\frac{\eta_{0i}}{\sqrt{T}} \right) \right)$$

$$= -\frac{1}{\sqrt{T}} \sum_{t=1}^{T} f_{it}(\kappa_{it}(\tau)) \left(\frac{\eta_{0i}}{\sqrt{T}} + \boldsymbol{x}_{it}' \frac{\boldsymbol{\eta}_1(\tau)}{\sqrt{TN}} \right) - \frac{\lambda_T}{\sqrt{T}} f_a(s_i) \frac{\eta_{0i}}{\sqrt{T}} + o(1)$$

Optimality of the $\hat{\eta}_{0i}$'s implies that $v(\eta_{0i}, \boldsymbol{\eta}_1) = o(T^{-1})$, and thus $\mathbb{E}(v(\eta_{0i}, \boldsymbol{\eta}_1) - v(0, \mathbf{0})) = v(0, \mathbf{0})$. Letting $f_i = T^{-1} \sum_{t=1}^{T} f_{it}(\kappa_{it}(\tau)) + \lambda_T/\sqrt{T} f_a(s_i)/\sqrt{T}$, we find that

$$\frac{\hat{\eta}_{0i}}{\sqrt{T}} = -(Tf_i)^{-1} \sum_{t=1}^{T} f_{it}(\kappa_{it}(\tau)) \boldsymbol{x}_{it}' \frac{\boldsymbol{\eta}_1(\tau)}{\sqrt{TN}} + \tilde{\boldsymbol{r}}_{it} \approx -\dot{\boldsymbol{x}}_i' \frac{\boldsymbol{\eta}_1(\tau)}{\sqrt{TN}}$$

where $\dot{\boldsymbol{x}}_i = \sum_t f_{it}(\kappa_{it}(\tau)) \boldsymbol{x}_{it}/Tf_i$. By Koenker's (2004) Theorem 1 and Lamarche's (2010) Lemma 1, the components of $\tilde{\boldsymbol{r}}$ are asymptotically negligible. Therefore, we replace $\hat{\eta}_{0i}/\sqrt{T}$ in the objective function, and we decompose the function in four parts:

$$V_{TN}^{(1)}(\boldsymbol{\eta}_1(\tau)) = -\sum_{t=1}^{T} \sum_{i=1}^{N} (\boldsymbol{x}_{it}' - \dot{\boldsymbol{x}}_i')(\boldsymbol{\eta}_1(\tau)/\sqrt{NT}) \psi_\tau(y_{it} - \kappa_{it}(\tau))$$

$$V_{TN}^{(2)}(\boldsymbol{\eta}_1(\tau)) = \sum_{t=1}^{T} \sum_{i=1}^{N} \int_0^{v_{it,TN}} (I(y_{it} - \kappa_{it}(\tau) \le s) - I(y_{it} - \kappa_{it}(\tau) \le 0)) ds$$

$$V_{TN}^{(3)}(\boldsymbol{\eta}_1(\tau)) = -\lambda_T \sum_{i=1}^{N} \dot{\boldsymbol{x}}_i' \left(\boldsymbol{\eta}_1(\tau)/\sqrt{NT} \right) \operatorname{sgn}(\alpha_i)$$

$$V_{TN}^{(4)}(\boldsymbol{\eta}_1(\tau)) = \lambda_T \sum_{i=1}^{N} \int_0^{\dot{\boldsymbol{x}}_i' \frac{\boldsymbol{\eta}_1(\tau)}{\sqrt{TN}}} (I(\alpha_i \le s) - I(\alpha_i \le 0)) ds$$

with $v_{it,TN} = (\boldsymbol{x}'_{it} - \dot{\boldsymbol{x}}'_i)\boldsymbol{\eta}_1(\tau)/\sqrt{TN}$. The first term is asymptotically Gaussian,

$$V_{TN}^{(1)}(\boldsymbol{\eta}_1(\tau)) = -\frac{1}{\sqrt{TN}} \sum_{t=1}^{T} \sum_{i=1}^{N} (\boldsymbol{x}'_{it} - \dot{\boldsymbol{x}}'_i)\boldsymbol{\eta}_1(\tau)\psi_\tau(y_{it} - \kappa_{it}(\tau)) \rightsquigarrow -\boldsymbol{\eta}_1(\tau)'\boldsymbol{B}$$

where $\boldsymbol{B}$ is a Gaussian vector with covariance $\boldsymbol{H}_0$. The second term converges in probability to a quadratic term in $\boldsymbol{\eta}_1(\tau)$,

$$\mathbb{E}V_{TN}^{(2)}(\boldsymbol{\eta}_1(\tau)) = \frac{1}{2TN} \sum_{t=1}^{T} \sum_{i=1}^{N} f_{it}(\kappa_{it}(\tau))((\boldsymbol{x}'_{it} - \dot{\boldsymbol{x}}'_i)\boldsymbol{H}_1(\tau))^2 + o(1) \to \frac{1}{2}\boldsymbol{\eta}_1(\tau)'\boldsymbol{H}_1\boldsymbol{\eta}_1(\tau)$$

The last two terms of $V_{TN}(\boldsymbol{\eta}_1(\tau))$ represents a decomposition of the stochastic penalty term. The third term is also asymptotically Gaussian,

$$V_{TN}^{(3)}(\boldsymbol{\eta}_1(\tau)) = -\frac{\lambda_T}{\sqrt{T}}\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \dot{\boldsymbol{x}}'_i\boldsymbol{\eta}_1(\tau)\text{sgn}(\alpha_i) \rightsquigarrow -\lambda\boldsymbol{\eta}_1(\tau)'\boldsymbol{C}$$

where $\boldsymbol{C}$ is a Gaussian vector independent of $\boldsymbol{B}$ with covariance $\boldsymbol{H}_2$. Lastly, the fourth term $V_{TN}^{(4)}(\boldsymbol{\eta}_1(\tau))$ is asymptotically quadratic in $\boldsymbol{\eta}_1(\tau)$,

$$\mathbb{E}V_{TN}^{(4)}(\boldsymbol{\eta}_1(\tau)) = \frac{\lambda_T}{2TN} \sum_{i=1}^{N} f_a(s_i)(\dot{\boldsymbol{x}}'_i\boldsymbol{\eta}_1(\tau))^2 + o(1) \to \frac{1}{2}\lambda\boldsymbol{\eta}_1(\tau)'\boldsymbol{H}_3\boldsymbol{\eta}_1(\tau)$$

Since $V_{TN}(\boldsymbol{\eta}_1(\tau))$ is convex and $V_0(\boldsymbol{\eta}_1(\tau))$ has a unique minimum, it follows that $\text{argmin}(V_{TN}(\bullet)) \rightsquigarrow \text{argmin}(V_0(\bullet))$. The penalized estimator converges to a Gaussian random variable with mean $(\boldsymbol{H}_1 + \lambda\boldsymbol{H}_3)^{-1}\lambda\mathbb{E}\boldsymbol{C}$ and covariance $\text{Avar}(\sqrt{NT}(\tilde{\boldsymbol{\beta}}(\tau,\lambda)) = (\boldsymbol{H}_1 + \lambda\boldsymbol{H}_3)^{-1}(\boldsymbol{H}_0 + \lambda^2\boldsymbol{H}_2)(\boldsymbol{H}_1 + \lambda\boldsymbol{H}_3)^{-1}$. Theorem 1 shows that $\sqrt{NT}(\hat{\boldsymbol{\beta}}(\tau,\lambda) - \boldsymbol{\beta}(\tau))$ converges in distribution to a Gaussian random variable with mean zero and covariance $\text{Avar}(\sqrt{NT}(\hat{\boldsymbol{\beta}}(\tau,\lambda)) = (\boldsymbol{\Sigma}_1 + \lambda\boldsymbol{\Sigma}_3)^{-1}(\boldsymbol{\Sigma}_0 + \lambda^2\boldsymbol{\Sigma}_2)(\boldsymbol{\Sigma}_1 + \lambda\boldsymbol{\Sigma}_3)^{-1}$. Lastly, we need to show that $\text{Avar}(\sqrt{NT}(\tilde{\boldsymbol{\beta}}(\tau,\lambda)) < \text{Avar}(\sqrt{NT}(\hat{\boldsymbol{\beta}}(\tau,\lambda))$.

Lemma 2 implies that $\boldsymbol{P}_{\boldsymbol{\Upsilon}}\boldsymbol{X}_o$ is equal to $\boldsymbol{P}_{\boldsymbol{\Phi}}\boldsymbol{X}_o$, we have that $\boldsymbol{H}_0 = \boldsymbol{\Sigma}_0 = \boldsymbol{J}_0$, $\boldsymbol{H}_2 = \boldsymbol{\Sigma}_2 = \boldsymbol{J}_2$, and $\boldsymbol{H}_3 = \boldsymbol{\Sigma}_3 = \boldsymbol{J}_3$. Notice that the conditional density of α_i at the median is equal to the unconditional density of a_i at zero. Therefore, the asymptotic relative efficiency between $\tilde{\boldsymbol{\beta}}(\tau,\lambda)$ and $\hat{\boldsymbol{\beta}}(\tau,\lambda)$ is determined by

$$\begin{aligned}\boldsymbol{H}_1 - \boldsymbol{\Sigma}_1 &= \boldsymbol{L}'\boldsymbol{\Phi}\boldsymbol{L} - \boldsymbol{R}'\boldsymbol{\Upsilon}\boldsymbol{R} = \boldsymbol{L}'(\boldsymbol{\Phi} - \boldsymbol{\Upsilon})\boldsymbol{L} \\ &= \boldsymbol{L}'\boldsymbol{\Phi}\boldsymbol{\Lambda}\boldsymbol{L} = ||\boldsymbol{L}||^2_{\boldsymbol{\Phi}\boldsymbol{\Lambda}} > 0,\end{aligned}$$

with the inequality indicating that $\boldsymbol{H}_1 - \boldsymbol{\Sigma}_1$ is positive definite and implying that the asymptotic variance of the penalized estimator $\tilde{\boldsymbol{\beta}}(\tau,\lambda)$ is smaller than the asymptotic variance of $\hat{\boldsymbol{\beta}}(\tau,\lambda)$ for all λ's in $\mathbb{R}_+$. $\qquad\square$

42

Proof of Corollary 1. Theorem 1 implies that the asymptotic mean squared error (AMSE) of $\hat{\boldsymbol{\beta}}(\tau, \lambda)$ is,

(A.1) $\qquad \text{trAMSE}(\hat{\boldsymbol{\beta}}(\tau, \lambda)) = \text{tr}\left\{(\boldsymbol{\Sigma}_1 + \lambda\boldsymbol{\Sigma}_3)^{-1}(\boldsymbol{\Sigma}_0 + \lambda^2\boldsymbol{\Sigma}_2)(\boldsymbol{\Sigma}_1 + \lambda\boldsymbol{\Sigma}_3)^{-1}\right\}.$

By Theorem 1, we can write A.1 as,

$$
\begin{aligned}
\text{trAMSE}(\hat{\boldsymbol{\beta}}(\tau, \lambda)) &= \text{tr}\left\{(\boldsymbol{\Sigma}_1 + \lambda\boldsymbol{J}_3)^{-1}(\boldsymbol{J}_0 + \lambda^2\boldsymbol{J}_2)(\boldsymbol{\Sigma}_1 + \lambda\boldsymbol{J}_3)^{-1}\right\} \\
&= \text{tr}\left\{(\boldsymbol{J}_3^{-1}\boldsymbol{\Sigma}_1 + \lambda\boldsymbol{I})^{-1}\boldsymbol{J}_3^{-1}\boldsymbol{J}_2(\boldsymbol{J}_2^{-1}\boldsymbol{J}_0 + \lambda^2\boldsymbol{I})(\boldsymbol{J}_3^{-1}\boldsymbol{\Sigma}_1 + \lambda\boldsymbol{I})^{-1}\boldsymbol{J}_3^{-1}\right\} \\
&= \text{tr}\left\{(\hat{\boldsymbol{B}} + \lambda\boldsymbol{I})^{-1}\boldsymbol{A}^{-1}\boldsymbol{C}(\boldsymbol{D} + \lambda^2\boldsymbol{I})(\boldsymbol{B} + \lambda\boldsymbol{I})^{-1}\boldsymbol{A}^{-1}\right\}.
\end{aligned}
$$

where $\boldsymbol{A} = \boldsymbol{J}_3$, $\hat{\boldsymbol{B}} = \boldsymbol{J}_3^{-1}\boldsymbol{\Sigma}_1$, $\boldsymbol{C} = \boldsymbol{J}_2$, and $\boldsymbol{D} = \boldsymbol{C}^{-1}\boldsymbol{J}_0$. By Lemma 1,

$$
\text{trAMSE}(\hat{\boldsymbol{\beta}}(\tau, \lambda)) = \sum_{i=1}^p \frac{\zeta_c^i(\zeta_d^i + \lambda^2)}{(\zeta_a^i(\zeta_{\hat{b}}^i + \lambda))^2}.
$$

Moreover, by Theorem 1, we write the AMSE of $\tilde{\boldsymbol{\beta}}(\tau, \lambda)$ as,

$$
\begin{aligned}
\text{trAMSE}(\tilde{\boldsymbol{\beta}}(\tau, \lambda)) &= \text{tr}\left\{(\boldsymbol{H}_1 + \lambda\boldsymbol{J}_3)^{-1}(\boldsymbol{J}_0 + \lambda^2\boldsymbol{J}_2)(\boldsymbol{H}_1 + \lambda\boldsymbol{J}_3)^{-1}\right\} \\
&\quad + \text{tr}\left\{\lambda^2((\boldsymbol{H}_1 + \lambda\boldsymbol{J}_3)^{-1}\boldsymbol{S}_o(\boldsymbol{H}_1 + \lambda\boldsymbol{J}_3)^{-1}\right\} \\
&= \text{tr}\left\{(\boldsymbol{J}_3^{-1}\boldsymbol{H}_1 + \lambda\boldsymbol{I})^{-1}\boldsymbol{J}_3^{-1}\boldsymbol{J}_2(\boldsymbol{J}_2^{-1}\boldsymbol{J}_0 + \lambda^2\boldsymbol{I})(\boldsymbol{J}_3^{-1}\boldsymbol{H}_1 + \lambda\boldsymbol{I})^{-1}\boldsymbol{J}_3^{-1}\right\} \\
&\quad + \text{tr}\left\{\lambda^2((\boldsymbol{J}_3^{-1}\boldsymbol{H}_1 + \lambda\boldsymbol{I})^{-1}\boldsymbol{J}_3^{-1}\boldsymbol{S}_o(\boldsymbol{J}_3^{-1}\boldsymbol{H}_1 + \lambda\boldsymbol{I})^{-1}\boldsymbol{J}_3^{-1}\right\} \\
&= \text{tr}\left\{(\tilde{\boldsymbol{B}} + \lambda\boldsymbol{I})^{-1}\boldsymbol{A}^{-1}\boldsymbol{C}(\boldsymbol{D} + \lambda^2\boldsymbol{I})(\tilde{\boldsymbol{B}} + \lambda\boldsymbol{I})^{-1}\boldsymbol{A}^{-1}\right\} \\
&\quad + \text{tr}\left\{\lambda^2(\tilde{\boldsymbol{B}} + \lambda\boldsymbol{I})^{-1}\boldsymbol{A}^{-1}\boldsymbol{S}_o((\tilde{\boldsymbol{B}} + \lambda\boldsymbol{I})^{-1}\boldsymbol{A}^{-1})'\right\}
\end{aligned}
$$

where $\boldsymbol{A} = \boldsymbol{J}_3$, $\tilde{\boldsymbol{B}} = \boldsymbol{J}_3^{-1}\boldsymbol{H}_1$, $\boldsymbol{C} = \boldsymbol{J}_2$, and $\boldsymbol{D} = \boldsymbol{C}^{-1}\boldsymbol{J}_0$. By Lemma 1,

$$
\text{trAMSE}(\tilde{\boldsymbol{\beta}}(\tau, \lambda)) = \sum_{i=1}^p \frac{\zeta_c^i(\zeta_d^i + \lambda^2)}{(\zeta_a^i(\zeta_{\tilde{b}}^i + \lambda))^2} + \frac{\bar{\zeta}_{S_o}\lambda^2}{(\bar{\zeta}_a(\bar{\zeta}_{\tilde{b}} + \lambda))^2}.
$$

$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Proof of Corollary 2. The trace of the normalized asymptotic covariance matrix of the penalized estimator is,

$$
\begin{aligned}
\text{trAVar}(\hat{\boldsymbol{\beta}}(\tau, \lambda)) &= \text{tr}\{(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}_1)(\boldsymbol{\Sigma}_1 + \lambda\boldsymbol{\Sigma}_3)^{-1}(\boldsymbol{\Sigma}_0 + \lambda^2\boldsymbol{\Sigma}_2)(\boldsymbol{\Sigma}_1 + \lambda\boldsymbol{\Sigma}_3)^{-1}\} \\
&= \text{tr}\left\{\boldsymbol{A}\boldsymbol{B}^{-1}\boldsymbol{A}(\boldsymbol{D} + \lambda\boldsymbol{I})^{-1}\boldsymbol{C}^{-1}\boldsymbol{E}(\boldsymbol{F} + \lambda^2\boldsymbol{I})(\boldsymbol{D} + \lambda\boldsymbol{I})^{-1}\boldsymbol{C}^{-1}\right\},
\end{aligned}
$$

where the matrices $\boldsymbol{A} = \boldsymbol{\Sigma}_1$, $\boldsymbol{B} = \boldsymbol{\Sigma}_0$, $\boldsymbol{C} = \boldsymbol{\Sigma}_3$, $\boldsymbol{D} = \boldsymbol{C}^{-1}\boldsymbol{A}$, $\boldsymbol{E} = \boldsymbol{\Sigma}_2$ and $\boldsymbol{F} = \boldsymbol{E}^{-1}\boldsymbol{B}$. Replacing the matrices by their spectral decomposition, we have that,

$$
\begin{aligned}
\text{trAVar}(\hat{\boldsymbol{\beta}}(\tau, \lambda)) &= \text{tr}\{\boldsymbol{U}_a\boldsymbol{\Lambda}_a\boldsymbol{U}_a'(\boldsymbol{U}_b\boldsymbol{\Lambda}_b\boldsymbol{U}_b')^{-1}\boldsymbol{U}_a\boldsymbol{\Lambda}_a\boldsymbol{U}_a'(\boldsymbol{U}_d\boldsymbol{\Lambda}_d\boldsymbol{U}_d' + \lambda\boldsymbol{I})^{-1}(\boldsymbol{U}_c\boldsymbol{\Lambda}_c\boldsymbol{U}_c')^{-1} \\
&\qquad \boldsymbol{U}_e\boldsymbol{\Lambda}_e\boldsymbol{U}_e'(\boldsymbol{U}_f\boldsymbol{\Lambda}_f\boldsymbol{U}_f' + \lambda^2\boldsymbol{I})(\boldsymbol{U}_d\boldsymbol{\Lambda}_d\boldsymbol{U}_d' + \lambda\boldsymbol{I})^{-1}(\boldsymbol{U}_c\boldsymbol{\Lambda}_c\boldsymbol{U}_c')^{-1}\} \\
&= \text{tr}\{\boldsymbol{\Lambda}_a\boldsymbol{\Lambda}_b^{-1}\boldsymbol{\Lambda}_a(\boldsymbol{\Lambda}_d + \lambda\boldsymbol{I})^{-1}\boldsymbol{\Lambda}_c^{-1}\boldsymbol{\Lambda}_e(\boldsymbol{\Lambda}_f + \lambda^2\boldsymbol{I})(\boldsymbol{\Lambda}_d + \lambda\boldsymbol{I})^{-1}\boldsymbol{\Lambda}_c^{-1}\} \\
&= \sum_{i=1}^{p} \frac{(\zeta_a^i)^2 \zeta_e^i(\zeta_f^i + \lambda^2)}{\zeta_b^i(\zeta_c^i(\zeta_d^i + \lambda))^2} = \sum_{i=1}^{p} \pi(\lambda)^i
\end{aligned}
$$

We now have a simple optimization problem as a function of λ, with positive eigenvalues ζ_k^i for all i, k. It then follows that the trace of the normalized asymptotic covariance matrix has a unique minimizer λ^* such that, $\text{trAVar}(\hat{\boldsymbol{\beta}}(\tau, \lambda^*)) < \text{trAVar}(\hat{\boldsymbol{\beta}}(\tau, \lambda))$. $\qquad\square$