Penalized Quantile Regression with Semiparametric Correlated Effects: An Application with Heterogeneous Preferences*

Matthew Harding† and Carlos Lamarche‡

January 24, 2015

Abstract

This paper proposes new ℓ_1-penalized quantile regression estimators for panel data, which explicitly allows for individual heterogeneity associated with covariates. Existing fixed effects estimators can potentially suffer from three limitations which are overcome by the proposed approach: (i) incidental parameters bias in non-linear models with large N and small T, (ii) lack of efficiency, and (iii) inability to estimate time-invariant effects. 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 technique to an empirical example of the estimation of consumer preferences for nutrients from a demand model using a large transaction level 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.

JEL: C21, C23, J22
Keywords: Shrinkage, Panel Data, Quantile Regression, Big Data

*The authors would like to thank Badi Baltagi, Erich Battistin, Zongwu Cai, Cheng Hsiao, Ted Juhl, Chihwa Kao, Roger Koenker, Roger Moon, Hashem Pesaran, Steve Portnoy, and Jim Ziliak for comments on a previous draft as well as seminar participants at the University of Southern California, University of Illinois at Urbana-Champaign, University of Kansas, University of Oklahoma, University of Padova, Syracuse University, the 17th Midwest Econometric Group conference, and the 25th meeting of the Latin American Econometric Society. We are also grateful to a Co-Editor and two anonymous referees for their comments which have improved the paper considerably.

†Corresponding author: Sanford School of Public Policy and Duke-UNC USDA Center for Behavioral Economics and Healthy Food Choice Research, Duke University, 140 Science Drive, Durham, NC 27708; Phone: (919) 613-1306; Fax: (919) 613-0539; Email: matthew.harding@duke.edu

‡Department of Economics, University of Kentucky, 335A Gatton College of Business and Economics, Lexington, KY 40506-0034. Tel.: +1 859-257-3371. Email: clamarche@uky.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 more recent developments, see Abrevaya and Dahl (2008), Graham, Hahn, and Powell (2009), Harding and Lamarche (2009, 2014a), Lamarche (2010), Galvao (2011), Canay (2011), 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 non-separable 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.

Flexibility in specification and unobserved heterogeneity has a long history of playing a fundamental role in the estimation of economic models (Burtless and Hausman (1978), Hausman (1985), Blundell and Meghir (1986), Ziliak and Kniesner (1999), Blundell and MaCurdy (1999), van Soest, Das, and Gong (2002), Kumar (2012), Dubois, Griffith and Nevo (2014), among others). 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 non-linear models with endogenous regressors. For instance, in empirical specifications for food expenditures, the quantity of nutrients purchased is suspected to be endogenous because unobserved time-invariant preferences for food may be correlated with latent factors associated with product attributes (see, e.g., Dubois, Griffith and Nevo (2014)).

The penalized quantile regression estimator can be extended to models with endogenous covariates. In addition to investigating this extension, this paper proposes a series of Hausman-type tests to evaluate exogeneity assumptions. 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)). Cai and Xiao (2012) study the estimation of a partially varying coefficient model in quantile regression which is estimated using semiparametric methods. Abrevaya and Dahl (2008) propose a quantile regression approach for a model that includes correlated random effects. 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)).[1] 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 effects, while simultaneously addressing issues associated with incidental parameters and correlation between independent variables and individual effects. Furthermore, it is not more difficult to implement than other quantile regression panel data methods.

One of the most significant contributions of this paper is the following aspect. 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

[1] Koenker and Hallock (2000) recognize the difficulties arising in panel quantile models due to the lack of a suitable transformation to remove unobserved heterogeneity potentially associated with covariates. Rosen (2012) points out to the possible lack of identification in models with additive error terms.

(2009) show that there is no incidental parameter problem in a non-differentiable panel data model and Koenker (2004) and Galvao, Lamarche, and 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 (2014) 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, and thus, it reduces the bias of the fixed effects panel data estimator which might arise from incidental parameters in small T panels.

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 provide an empirical example which addresses the problem of estimating preference heterogeneity in consumer demand models using large transaction level 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:

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

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

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 $\boldsymbol{\beta}(\tau)$ models the covariate effect providing an opportunity for investigating how 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 2.2 and 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:

$$\alpha_i(\boldsymbol{x}_i, u_{i1}, a_i) = \alpha_i(\boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}, ..., \boldsymbol{x}_{iT}, u_{i1}, a_i) = g(\boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}, ..., \boldsymbol{x}_{iT}, u_{i1}) + a_i, \tag{2.3}$$

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 monotonicity 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 static version of their model arises by replacing $\boldsymbol{x}_i$ by t_i in equation (2.3). The conditional quantile function is equal to $Q_{Y_{it}}(\tau|t_i, \boldsymbol{x}_{it}, a_i) = g(\tau, t_i) + \boldsymbol{x}_{it}'\boldsymbol{\beta}(\tau) + a_i$.

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,

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

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 degree of 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,

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

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 $\|\boldsymbol{a}\|_1 := \sum_i |a_i|$.

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., Horowitz and Lee (2007), Carrasco, Florens and Renault (2007), Chen (2007), Belloni and Chernozhukov (2011), Belloni, Chen, Chernozhukov and Hansen (2012); see also Bickel and Li (2006) for a survey in statistics).

Although flexibility in specification is naturally important, the estimation of econometric models is typically 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:

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

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.

In the case of the one-step estimator, the method considers the following design matrix,

$$\begin{bmatrix} \text{diag}(\omega) \otimes \boldsymbol{X} & \text{diag}(\omega) \otimes \boldsymbol{ZD} & \omega \otimes \boldsymbol{Z} \\ \boldsymbol{0} & \boldsymbol{0} & \lambda \boldsymbol{I} \end{bmatrix},$$

where $\boldsymbol{X}$ is a $NT \times p$ matrix, $\boldsymbol{D}$ is a $NT \times pT$ matrix, $\boldsymbol{Z}$ is a $NT \times N$ matrix and

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

In the application considered in Section 5, we show that it is also possible to estimate time-invariant effects by introducing additional explanatory variables in $\boldsymbol{X}$. In this case, for identification when λ is exactly zero, we will need the standard conditions on restricting the estimation to $N-1$ individual specific effects. Note that in a model with time-invariant variables perfect collinearity arises in the limiting case where λ is equal to zero, which is not considered in this paper. Our estimator is defined for $\lambda > 0$, although λ can be very small. When $\lambda > 0$, there are $\hat{a}_i$'s that are exactly zero, which is equivalent to a model with $m < N$ individual effects. This allows identification of time-invariant effects and does not impose biases under the assumption that a_i is independent of the covariates.

2.2. **A simple extension**

A natural implicit restriction of the framework is that time-invariant variables are assumed to be exogenous. This section proposes a way out of this problem motivated by several approaches for correlated random effects models (e.g., Arellano and Bover (1995) and Ziliak (2003) for an implementation of the estimator). Consider for example a model with time-varying and time-invariant independent variables:

$$y_{it} = \boldsymbol{x}'_{it}\boldsymbol{\beta} + \boldsymbol{w}'_i\boldsymbol{\delta} + \alpha_i + u_{it}, \tag{2.7}$$

where y_{it} is the response variable for subject i at time t, $\boldsymbol{x}_{it} = (\boldsymbol{x}_{1,it}, \boldsymbol{x}_{2,it})'$, $\boldsymbol{w}_i = (\boldsymbol{w}'_{1,i}, \boldsymbol{w}'_{2,i})'$, and u_{it} is the error term. The variable $\boldsymbol{x}_{1,it}$ is a p_1-dimensional vector of endogenous variables, $\boldsymbol{x}_{2,it}$ is a p_2-dimensional vector of exogenous independent variables that includes an intercept. Similarly, the variable $\boldsymbol{w}_{1,i}$ is a k_1-dimensional vector of endogenous variables, and $\boldsymbol{w}_{2,i}$ is a k_2-dimensional vector of exogenous independent variables. Following (2.7) and the representation for α_i in equation (2.3), we have that, in the case of a linear model $\alpha_i =$

$\boldsymbol{x}_i'\boldsymbol{\gamma} + a_i$, where as before $\boldsymbol{x}_i'\boldsymbol{\gamma} = \sum_{t=1}^{T} \boldsymbol{x}_{it}'\boldsymbol{\gamma}_t$. The difference is now that we allow a_i to be correlated with $\boldsymbol{w}_{1i}$.

This model can be estimated using an extension of the penalized estimator which employs instrumental variables (see, e.g., Harding and Lamarche (2014b)). For instance, in the case of the one-step estimator, we consider the following objective function:

$$(2.8)\quad \sum_{j=1}^{J}\sum_{t=1}^{T}\sum_{i=1}^{N} \omega_j \rho_{\tau_j}\left(y_{it} - \boldsymbol{x}_{it}'\boldsymbol{\beta}(\tau_j) - \boldsymbol{w}_i'\boldsymbol{\delta}(\tau_j) - \sum_{k=1}^{T} \boldsymbol{x}_{ik}'\boldsymbol{\gamma}_k(\tau_j) - a_i - \boldsymbol{\phi}_i'\boldsymbol{\pi}(\tau_j)\right) + \lambda\|\boldsymbol{a}\|_1,$$

where τ_j the quantile of interest and $\boldsymbol{\phi}_i$ is a vector of instruments. Note that we address the endogeneity of the time-variant variables by augmenting the design matrix as before. However, the individual effects a_i's can be correlated with the endogenous time-invariant variables. Provided that $k_1 \leq p_2$, we can follow Hausman and Taylor (1981) and construct instruments using transformations of the exogenous variables. For instance, the time-invariant endogenous variable can be instrumented using a vector of individual specific variables $\bar{\boldsymbol{x}}_{2,i} = T^{-1}\sum_{t=1}^{T} \boldsymbol{x}_{2,it}$.[2]

The proposed approach follows two steps. For a given λ, we minimize (2.8) over $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, $\boldsymbol{\delta}_2$, $\boldsymbol{a}$, and $\boldsymbol{\pi}$ as functions of $\boldsymbol{\tau}$ and $\boldsymbol{\delta}_1$. Then we estimate the coefficient on the endogenous time-invariant variable by finding the value of $\boldsymbol{\delta}_1(\boldsymbol{\tau})$, which minimizes a weighted distance function defined on $\boldsymbol{\pi}$:

$$(2.9)\qquad \hat{\boldsymbol{\delta}}_1(\boldsymbol{\tau}) = \underset{\boldsymbol{\pi}_1\in\Pi_1}{\operatorname{argmin}}\ \hat{\boldsymbol{\pi}}(\boldsymbol{\tau},\boldsymbol{\delta}_1)'\boldsymbol{A}\hat{\boldsymbol{\pi}}(\boldsymbol{\tau},\boldsymbol{\delta}_1),$$

for a given positive definite matrix $\boldsymbol{A}$. Then, we obtain a two-step instrumental variable estimator defined as $(\hat{\boldsymbol{\beta}}(\boldsymbol{\tau},\hat{\boldsymbol{\delta}}_1(\boldsymbol{\tau}))', \hat{\gamma}(\boldsymbol{\tau},\hat{\boldsymbol{\delta}}_1(\boldsymbol{\tau}))', \hat{\boldsymbol{\delta}}_1(\boldsymbol{\tau})', \hat{\boldsymbol{\delta}}_2(\boldsymbol{\tau},\hat{\boldsymbol{\delta}}_1(\boldsymbol{\tau}))', \hat{\boldsymbol{a}}(\hat{\boldsymbol{\delta}}_1(\boldsymbol{\tau}))')'$. As an alternative to the approach proposed in (2.8) and (2.9), one can accomodate the recent control variate approach proposed by Chernozhukov, Fernández-Val and Kowalski (2015) to the case of endogenous continuous time-invariant variables.

2.3. **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. As shown in Section 5.2, the family of estimates associated with the slope coefficients lies on a one-dimensional path of finite length

[2]It is important to clarify that as in Harding and Lamarche (2009), instruments can be defined outside the model (e.g., the practitioner does not need to satisfy $k_1 \leq p_2$ as long as valid instrumental variables are available).

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, $\hat{\lambda} = \arg\inf \|\hat{u}(\tau,\lambda)\|_1 + \mathrm{df}_\lambda/(2NT)$, 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, $\hat{\lambda} = \arg\inf \{tr\boldsymbol{\Sigma}_{\boldsymbol{\beta}}\}$, where $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ 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 $\hat{\boldsymbol{\beta}}(\tau,0)$ 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 $\boldsymbol{s}_{0,it} = (\mathbb{E}(\mathrm{sign}(\alpha_i)\boldsymbol{x}_{it}))_{it}$ and $\boldsymbol{S}_0 = \boldsymbol{s}_0\boldsymbol{s}_0'$. The largest eigenvalue is defined as $\bar{\zeta}_{S_o} = \max\{\zeta^1_{S_o}, \ldots, \zeta^p_{S_o}\}$, 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,*

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

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:*

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

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 $\hat{\boldsymbol{\beta}}(\tau, \lambda)$ implies choosing λ to minimize asymptotic variance. The primary objective is now to show that the trace of the asymptotic covariance matrix of $\hat{\boldsymbol{\beta}}(\tau, \lambda)$ 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 $\text{Avar}(\hat{\beta}_k(\tau,\lambda))/\text{Avar}(\hat{\beta}_k(\tau,0))$.

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

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 2.3.

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. Finally, 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.

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 two 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 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 reports 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 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

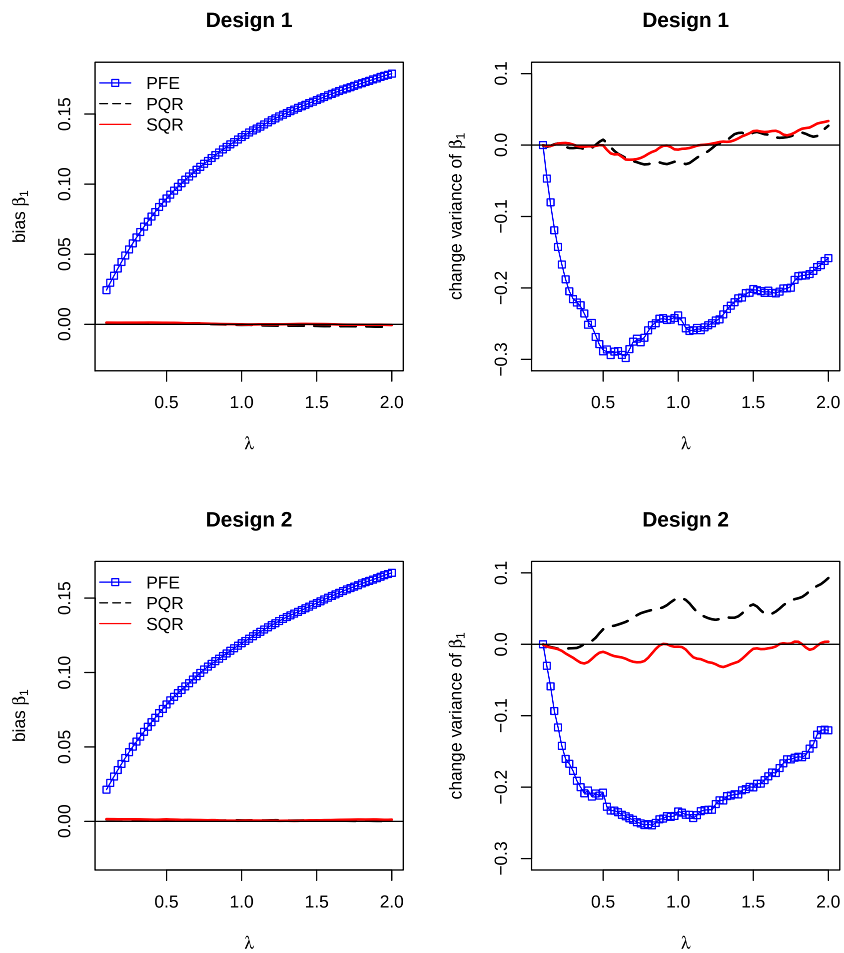

FIGURE 4.1. *Small sample performance of penalized quantile regression estimator for $\beta_1(0.5)$. The left panel shows the bias of the estimator and the right panel shows the variance percentage change.*

(ii) there are significant differences in variance reduction. The variance of the slope PFE estimator for λ around 0.5 is reduced more than 25 percent relative to the variance of the PFE estimator for $\lambda \to 0$. In contrast, the variance of the PQR estimator for $\lambda = 1$ relative to the PQR estimator for $\lambda \to 0$ is reduced by 2 percent.[3]

The PFE 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

[3]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.

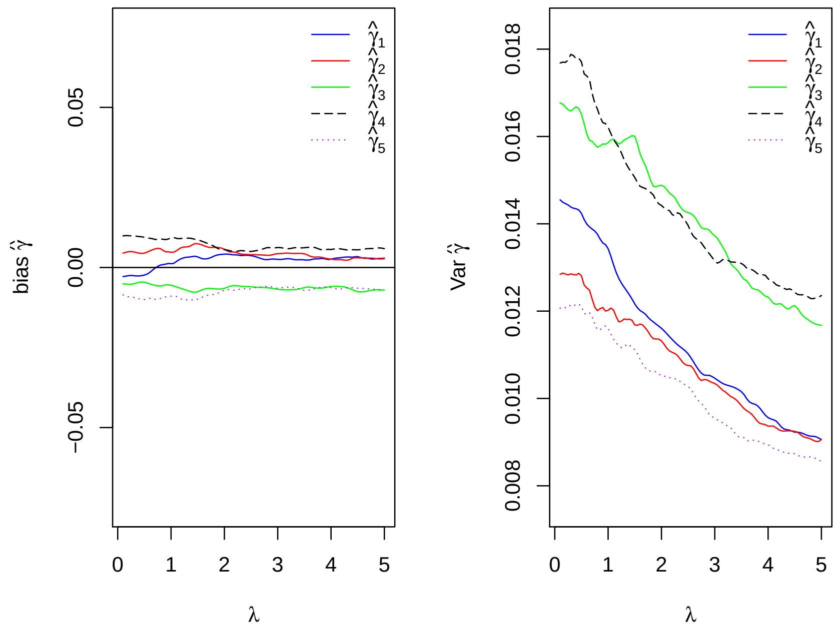

FIGURE 4.2. *Small sample performance of the penalized quantile regression estimator PQR for* $\boldsymbol{\gamma}$.

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, SQR, 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 we increase λ reflecting potential significant improvements in the performance of the estimator that penalizes exogenous individual effects.

Using Table 4.1, we report evidence of the performance of the methods under Design 3 considering several sample sizes $N = \{100, 500\}$ and $T = \{2, 12\}$. 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 produce unbiased results at the 0.5 quantile. However, in the location-scale shift

N	T		Least Squares		Quantile Regression						
			OLS	GLS	QR	PFE	CQR	2SQR	SQR	PQRd	PQR
						Design 3: $\tau = 0.50$					
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 3: $\tau = 0.25$					
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.*

model at $\tau = 0.25$, the fixed effects estimator 2SQR is significantly biased and PQR and SQR estimators offer the best overall performance related to existing methods. This result is expected because fixed effects methods, although they address the correlation between α_i and x_{it}, create incidental parameters bias when T is small.[4,5]

[4]Additional monte carlo evidence, not reported in this paper to avoid repetition and save space, shows that the fixed effects quantile regression estimator (FEQR) does not perform well in these simulations with biases similar to the 2SQR estimator. The results are available upon request.

[5]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 investigated the performance of PQR estimator by estimating λ^* using (i) the asymptotic covariance matrix and (ii) the bootstrap. The results suggested that there are no important efficiency losses when the practitioner estimates λ^*, at least in the models considered in this study. We found that the PQR estimator improves on the performance of the CQR estimator, offering considerable efficiency gains in all variants of the model.

5. 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 a complex problem that requires us to tailor interventions to account for heterogeneous preferences and demographics. In this example, we explore how preferences for nutrients vary by gender across the conditional food expenditure distribution.[6]

A recent report by the Trust for America's Health and the Robert Wood Johnson Foundation finds that while obesity rates appear to be stabilizing, the rates nevertheless continue to be very high.[7] More than two thirds of adults are now obese or overweight, and almost 35% of adults are obese. At the same time, a recent CDC report estimates that while 35.7% of US adults are now obese, substantial differences exist across race and gender groups.[8] 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%. Furthermore, while 56.6% of African American women, and 37.1% of African American men are obese, only 32.8% of White women and 32.4% of White men are obese. In this section, we quantify preferences for nutrients by gender across the conditional food expenditure distribution. Understanding the heterogeneity in preferences is of major policy interest as it helps us design better policies by accounting for their distributional impact. Distributional considerations play an important role in the design of US policies. Recently, Berkeley passed the first per-ounce soda tax on sugar-sweetened beverages in the US. Similar taxes have already been implemented in a number of other countries including Mexico. Distributional considerations of these taxes are being vigorously debated by both supporters and opponents of the taxes.

The consumption model described below allows us to quantify the differences in preferences for nutrients across sub-populations. Dubois, Griffith, and Nevo (2014) use this modeling

---

[6] A previous version of the paper included an application of the new method to the estimation of labor supply models. We demonstrated how the penalized estimator can be obtained and used to estimate quantile specific elasticities. The interested reader can find additional details in Harding and Lamarche (2013).

[7] See Levi, Segal, St. Laurent, and Rayburn (2014) for an in-depth analysis based on data from the 2011-2012 Behavioral Risk Factor Surveillance System database.

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

approach to show that there are substantial differences across countries. We adapt their 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 M 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 which are salient to the average consumer, C, 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 $\varkappa_k = \{\varkappa_{k,1}, \ldots, \varkappa_{k_C}\}'$ units of the underlying product attributes.[9]

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 (2014) and assume that the household derives utility from both the goods purchased and the underlying quantities $\boldsymbol{x}_i$ 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 available for purchase. Thus,

$$\max_{X_{i0}, \boldsymbol{X}_i} U(X_{i0}, \boldsymbol{X}_i, \boldsymbol{x}_i, \mu_i), \quad \text{s.t.} \quad X_{i0} + \boldsymbol{p}_i' \boldsymbol{X}_i = I_i, \tag{5.1}$$

where $x_{ik} = \varkappa_k' \boldsymbol{X}_i$, 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 $\boldsymbol{x}_i$, 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 (2014) 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.

[9] 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.

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 (2014) 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:

$$U(X_{i0}, \boldsymbol{X}_i, \boldsymbol{x}_i, \mu_i) = \exp(X_{i0}) \left(\sum_{k=1}^{D} f_{ik}(X_{ik}) \right)^{\mu_i} \prod_{c=1}^{C} h_{ic}(x_{ic}), \tag{5.2}$$

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 $h_{ic}(x_{ic}) = \exp(\beta_c x_{ic})$. After substituting for the budget constraint and the home production function the household maximizes the following log utility function:

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

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:

$$p_{ik} = \mu_i \theta_i \frac{\lambda_{ik} X_{ik}^{\theta_i - 1}}{\sum_{k=1}^{D} \lambda_{ik} X_{ik}^{\theta_i}} + \sum_{c=1}^{C} \beta_c \varkappa_{kc}. \tag{5.3}$$

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

$$p_{ik} X_{ik} = \mu_i \theta_i \frac{\lambda_{ik} X_{ik}^{\theta_i}}{\sum_{k=1}^{D} \lambda_{ik} X_{ik}^{\theta_i}} + \sum_{c=1}^{C} \beta_c \varkappa_{kc} X_{ik}. \tag{5.4}$$

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

$$E_i = \mu_i \theta_i + \sum_{c=1}^{C} \beta_c x_{ic}. \tag{5.5}$$

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

$$E_{it} = \alpha_i + \boldsymbol{w}_i' \boldsymbol{\delta} + \sum_{c=1}^{C} \beta_c x_{itc} + u_{it}, \tag{5.6}$$

where a latent term α_i and a vector of time invariant household demographics $\boldsymbol{w}_i$ is introduced to approximate household specific parameters $\mu_i\theta_i$. The error term u_{it} is an i.i.d. random variable 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:

$$\alpha_i = g(\boldsymbol{x}_i) + a_i, \tag{5.7}$$

where $g(\cdot)$ is an unknown function capturing household specific nutrition effects, $\boldsymbol{x}_i = (x_{i11}, \ldots, x_{i1C}, x_{i21}, \ldots, x_{i2C}, \ldots, x_{iTC})'$ captures the vector of time-varying values of the observed covariates and a_i is a household specific effect, potentially uncorrelated with the RHS variables. We investigate this condition in Section 5.3.

We estimate this model using the following quantile representation,

$$E_{it} = \sum_{c=1}^{C} \beta_c(u_{it}) x_{itc} + \boldsymbol{w}_i' \boldsymbol{\delta}(u_{it}) + g(\boldsymbol{x}_i, u_{it}) + a_i \tag{5.8}$$

$$\tau \mapsto \sum_{c=1}^{C} \beta_c(\tau) x_{itc} + \boldsymbol{w}_i' \boldsymbol{\delta}(\tau) + g(\boldsymbol{x}_i, \tau) + a_i \tag{5.9}$$

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.

5.1. **Data description**

We use a subset of the data introduced in Harding and Lovenheim (2014), 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. Participants in the panel are required to re-scan each purchased product at home. Each week the scanned data is transmitted to the company where the household data is combined with and verified against store sale information. As a result the data contains fairly 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.

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 (milligrams 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 5.1. *Descriptive statistics for the Nielsen HomeScan sample of single households in 2010. The table presents the sample mean and standard deviation (in parenthesis).*

While providing detailed information on each transaction, the Nielsen data does not record the nutritional content of every product. Harding and Lovenheim (2014) 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 5.1). 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.

	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)	(0.053)	(0.051)	(0.028)	(0.063)	(0.068)	(0.058)	(0.067)	(0.069)
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)	(0.043)	(0.043)	(0.021)	(0.029)	(0.030)	(0.044)	(0.035)	(0.039)
Cholesterol	0.091	0.177	0.175	-0.045	0.061	0.064	-0.145	0.008	0.007
	(0.011)	(0.024)	(0.025)	(0.010)	(0.024)	(0.024)	(0.016)	(0.018)	(0.019)
Protein	-0.013	-0.120	-0.115	2.761	1.856	1.842	5.134	2.960	3.020
	(0.110)	(0.245)	(0.263)	(0.138)	(0.346)	(0.365)	(0.222)	(0.283)	(0.286)
Carbohydrates	0.414	0.552	0.542	0.356	0.437	0.438	0.331	0.388	0.406
	(0.017)	(0.029)	(0.030)	(0.019)	(0.035)	(0.036)	(0.033)	(0.030)	(0.031)
Black	-5.305		-6.135	-6.717		-4.700	-8.570		-3.893
	(0.427)		(1.128)	(0.410)		(1.163)	(0.884)		(1.338)
High income	6.368		7.134	13.990		12.915	22.152		20.305
	(0.561)		(1.147)	(0.539)		(1.252)	(0.974)		(1.646)
College education	3.218		3.979	6.421		5.761	10.707		8.468
	(0.319)		(0.748)	(0.316)		(0.880)	(0.653)		(1.047)

TABLE 5.2. *Estimating a panel quantile model for female consumers. QR refers to quantile regression, FEQR is fixed effects quantile regression and SQR is the penalized estimator for the semiparametric correlated random effects model.*

5.2. **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 5.2, we report the estimated coefficients for the QR, FEQR and SQR estimators 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 5.1 and 5.2. 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 5.1, we report the estimated preferences and time-invariant coefficients for females, while Figure 5.2 reports the difference between the estimated coefficients for females and males.

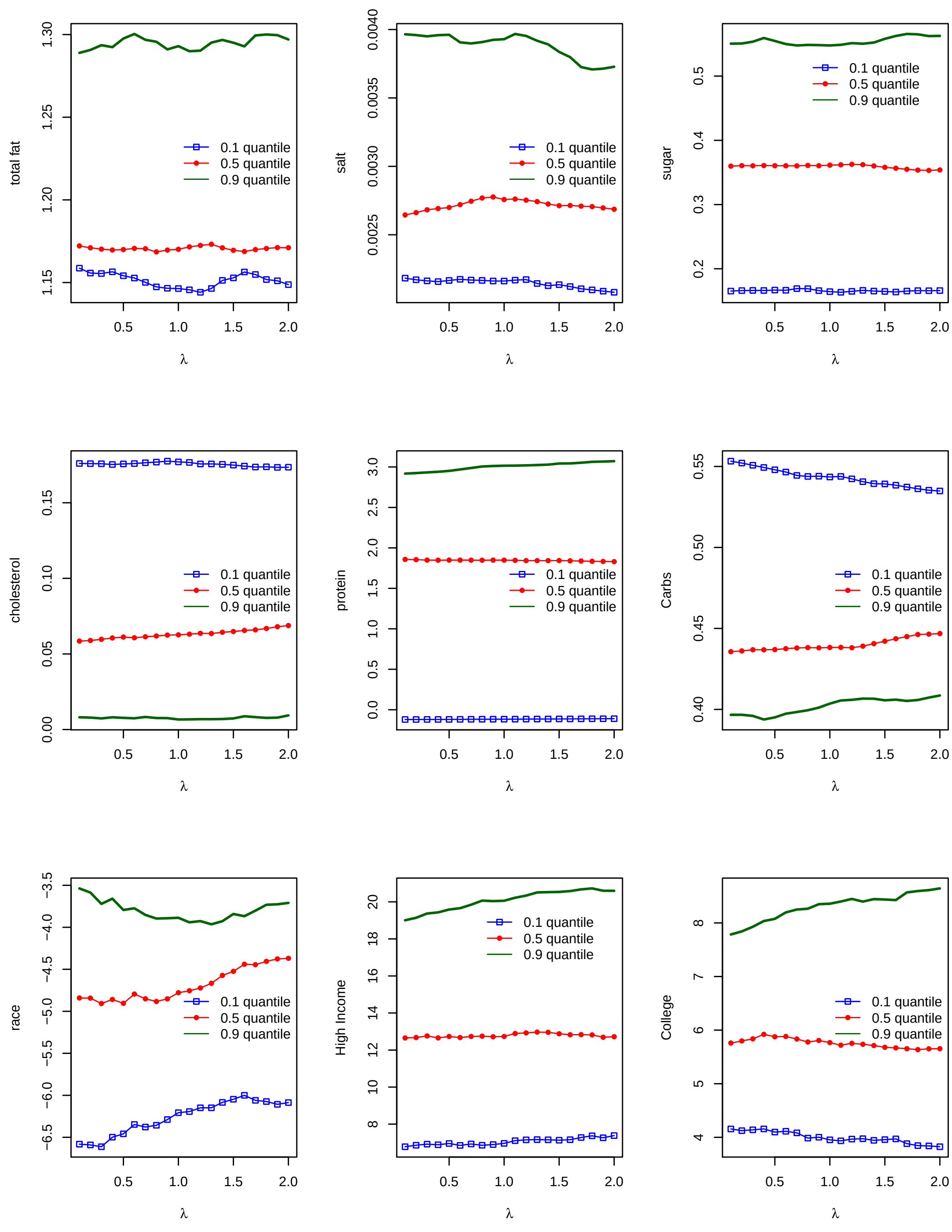

FIGURE 5.1. *Profile of preferences for female consumers.*

First, consider the results for female shoppers. We find that higher conditional food expenditures are associated with stronger preferences for fat, salt, sugar, and protein at the 0.5 and 0.9 quantiles. 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. At the same time taste matters a lot and drives consumer choices towards products which have ingredients that make them more appealing to the palate. As a result diets which are predicated on the idea of reducing the intake of one particular nutrients are not necessarily healthier overall, since they lead to increased consumption of other nutrients with overall ambiguous health outcomes.

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 insignificant at the 0.1 quantile, while positive and significant at the 0.9 quantile. Furthermore, our approach allows us to 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. To correctly interpret these demographic gradients it is important to recall that the data in this sample only captures food purchased in stores for at-home consumption. Lower expenditures are thus also associated with a higher propensity to eat outside the home at unhealthy fast-food outlets. Note that the results of the SQR model are quantitatively similar to those of the FEQR model, but substantially different than the QR results. This reinforces 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.

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 effect a_i is not centered at

	τ	SQR estimate at λ equal to					SQR: $\lambda = 0.01$ with			
		0.01	0.5	1.0	1.5	2.0	0.5	1.0	1.5	2.0
Total Fat	0.1	1.16	1.15	1.15	1.15	1.15	1.00	0.97	0.99	0.97
	0.5	1.17	1.17	1.17	1.17	1.17	1.00	1.00	1.00	1.00
	0.9	1.29	1.30	1.29	1.30	1.30	0.99	1.00	1.00	0.99
Salt ($\times$ 100)	0.1	0.22	0.22	0.22	0.21	0.21	1.00	1.00	0.98	0.92
	0.5	0.26	0.27	0.28	0.27	0.27	0.98	0.93	0.98	0.99
	0.9	0.40	0.40	0.39	0.38	0.37	1.00	1.00	0.99	0.89
Sugar	0.1	0.17	0.17	0.16	0.16	0.17	1.00	1.00	1.00	1.00
	0.5	0.36	0.36	0.36	0.36	0.35	1.00	1.00	1.00	0.99
	0.9	0.55	0.55	0.55	0.56	0.56	1.00	1.00	0.99	0.98
Cholesterol	0.1	0.18	0.18	0.18	0.17	0.17	1.00	1.00	1.00	0.99
	0.5	0.06	0.06	0.06	0.06	0.07	0.99	0.99	0.97	0.93
	0.9	0.01	0.01	0.01	0.01	0.01	1.00	1.00	1.00	1.00
Protein	0.1	-0.12	-0.12	-0.12	-0.11	-0.11	1.00	1.00	1.00	1.00
	0.5	1.86	1.85	1.85	1.84	1.83	1.00	1.00	1.00	1.00
	0.9	2.92	2.95	3.02	3.04	3.07	0.99	0.96	0.93	0.91
Carbs	0.1	0.55	0.55	0.54	0.54	0.53	0.98	0.95	0.89	0.79
	0.5	0.44	0.44	0.44	0.44	0.45	1.00	1.00	0.99	0.96
	0.9	0.40	0.40	0.40	0.41	0.41	1.00	0.99	0.97	0.95
Black	0.1	-6.58	-6.46	-6.21	-6.05	-6.09	0.99	0.93	0.83	0.86
	0.5	-4.84	-4.90	-4.78	-4.52	-4.37	1.00	1.00	0.95	0.88
	0.9	-3.54	-3.79	-3.89	-3.84	-3.71	0.98	0.96	0.97	0.99
High Income	0.1	6.78	6.96	6.96	7.13	7.38	0.99	1.00	0.96	0.93
	0.5	12.65	12.73	12.73	12.88	12.72	1.00	1.00	0.99	1.00
	0.9	19.01	19.59	20.06	20.54	20.60	0.95	0.92	0.82	0.84
College Education	0.1	4.16	4.10	3.95	3.96	3.82	1.00	0.97	0.97	0.92
	0.5	5.76	5.88	5.77	5.68	5.65	0.99	1.00	0.99	0.99
	0.9	7.78	8.08	8.36	8.44	8.64	0.92	0.82	0.81	0.69

Table 5.3. *Specification tests in a quantile regression model with semiparametric correlated effects. The last columns present the p-values of the tests.*

zero, this indicates that the model retains additional unobserved variables 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 at the 0.1 quantile, and income and college education at the 0.9 quantile (Figure 5.1). Next section develops a series of Hausman-type tests (Hausman 1978) to formally address these questions.

	τ	PQR estimate at λ equal to					PQR: $\lambda = 0.01$ with			
		0.01	0.5	1.0	1.5	2.0	0.5	1.0	1.5	2.0
Total Fat	0.1	1.16	1.16	1.15	1.15	1.16	1.00	0.96	0.97	0.98
	0.5	1.17	1.16	1.15	1.15	1.14	0.96	0.89	0.78	0.64
	0.9	1.29	1.29	1.28	1.27	1.28	1.00	0.95	0.87	1.00
Salt ($\times$ 100)	0.1	0.22	0.21	0.21	0.20	0.21	0.99	0.94	0.92	0.95
	0.5	0.27	0.27	0.28	0.28	0.27	0.97	0.92	0.94	0.94
	0.9	0.40	0.40	0.40	0.39	0.38	1.00	1.00	0.99	0.98
Sugar	0.1	0.17	0.16	0.17	0.17	0.17	1.00	1.00	1.00	0.99
	0.5	0.36	0.36	0.37	0.37	0.36	0.99	0.99	0.99	1.00
	0.9	0.55	0.56	0.56	0.57	0.58	0.97	0.98	0.94	0.86
Cholesterol	0.1	0.18	0.18	0.18	0.18	0.17	1.00	0.99	0.99	0.78
	0.5	0.06	0.06	0.06	0.06	0.06	1.00	1.00	1.00	1.00
	0.9	0.01	0.00	0.00	0.00	0.00	0.93	0.82	0.78	0.78
Protein	0.1	-0.12	-0.12	-0.12	-0.12	-0.12	1.00	1.00	1.00	1.00
	0.5	1.86	1.90	1.94	1.97	2.00	0.98	0.94	0.85	0.61
	0.9	2.93	3.01	3.11	3.18	3.26	0.79	0.52	0.23	0.02
Carbs	0.1	0.55	0.55	0.54	0.54	0.54	0.94	0.68	0.64	0.23
	0.5	0.43	0.43	0.43	0.43	0.43	0.98	0.98	0.98	0.98
	0.9	0.40	0.39	0.39	0.39	0.39	0.85	0.97	0.95	0.76
Black	0.1	-7.90	-7.85	-7.63	-7.60	-7.67	0.99	0.83	0.82	0.91
	0.5	-6.14	-6.12	-6.18	-6.02	-5.95	1.00	1.00	0.97	0.93
	0.9	-4.82	-5.06	-5.34	-5.61	-5.47	0.90	0.71	0.51	0.63
High Income	0.1	7.09	7.03	6.89	6.84	7.06	1.00	0.99	0.98	1.00
	0.5	12.91	12.79	12.68	12.45	12.23	1.00	0.99	0.94	0.91
	0.9	19.27	19.46	20.05	20.15	20.01	0.99	0.90	0.91	0.96
College Education	0.1	4.16	4.10	4.08	4.12	4.17	0.98	0.98	1.00	1.00
	0.5	5.77	5.82	5.78	5.83	5.88	0.98	1.00	0.99	0.96
	0.9	7.79	8.03	8.46	8.72	9.11	0.78	0.33	0.16	0.03

TABLE 5.4. *Specification tests in a quantile regression model with correlated effects. The last columns present the p-values of the tests.*

5.3. **Hypothesis testing**

To investigate if the model specification and assumptions are supported by data, we propose a series of tests. Recall that $\boldsymbol{x}_{it}$ denotes time-variant variables and $\boldsymbol{w}_i$ denotes time-invariant variables as in the model description. If we let $\alpha_i = g(\boldsymbol{x}_i) + a_i$ and α_i is correlated with $\boldsymbol{w}_i$, then a_i must be correlated with $\boldsymbol{w}_i$. Therefore, Assumption A2 in Appendix A.2 is violated, and shrinking a_i towards zero leads to biased estimates of time-invariant effects. On the other hand, the performance of the fixed effects version of the estimator (the limiting case of the penalized estimator when $\lambda \to 0$) does not depend on the correlation between a_i and $\boldsymbol{w}_i$.

It seems natural then to propose a Hausman-type test to evaluate whether the fixed effects estimator, $\lim_{\lambda \to 0} \hat{\beta}(\lambda)$ and the penalized estimator $\hat{\beta}(\lambda)$ offer significantly different results. If a_i and $\boldsymbol{w}_i$ are independent, we expect that $\lim_{\lambda \to 0} \hat{\beta}(\lambda)$ and $\hat{\beta}(\lambda)$ be relatively similar for any value of λ.

Tables 5.3 and 5.4 show point estimates and p-values of tests associated with the significance of difference between fixed effects and penalized quantile regression results. The null hypothesis is equality of coefficients and we evaluate several null hypotheses considering different values of $\lambda = \{0.5, 1, 1.5, 2\}$. These values are chosen for convenience and represent points on the profile of the penalized estimator in terms of λ (see, e.g., Figure 5.1). The specification tests are performed on two models. First, we estimate equation (5.9) using the semiparametric estimator, SQR. The results are shown in Table 5.3. We also estimate the model using the one-step estimator, PQR, and compare the results to emphasize the role of the semiparametric specification of the correlated effects (Table 5.4).

The results in Table 5.3 indicate that preference for nutrients do not significantly change when the shrinkage parameter changes. This result might seem clear from observing the point estimates presented in the first columns and the p-values of the tests which naturally offer more reliable evidence. We interpret these findings as suggesting the empirical findings are robust to the potential misspecification of equation (5.7). It is interesting to compare the result with Table 5.4 and observe that the semiparametric specification of the nutrients helps to correct for potential issues of endogeneity in the one-step estimator. For instance, the effects of protein at the 0.5 and 0.9 quantiles increase with λ if the one-step estimator is employed, while they appear to be fairly constant if the two-step semiparametric estimator is used. The result suggests that the semiparametric specification of the correlated effects helps address potential endogeneity issues. The results continue to provide support in favor of a correctly specified model for the time-invariant variables.[10] Overall, it is important to emphasize that the results support the penalized effects specification with semiparametric correlated effects and this model is not rejected in favor of the fixed effects specification. This is true specifically among preferences for nutrients among female and male consumers.[11]

[10]The results for college education in a model with parametric correlated effects estimated at the 0.9 quantile reveal some potential endogeneity issues. This naturally suggests that the effect of education at the upper tail cannot be interpreted as a causal effect. As a referee suggested, the effect might be interpreted as the sum of the causal effect of education on expenditures at the upper tail of the conditional distribution plus the effect of college education on time-invariant preferences.

[11]We repeat the exercise of estimating equation (5.9) considering a sample of male consumers. The results, which are not presented in the paper to economize space, continue to support the semiparametric

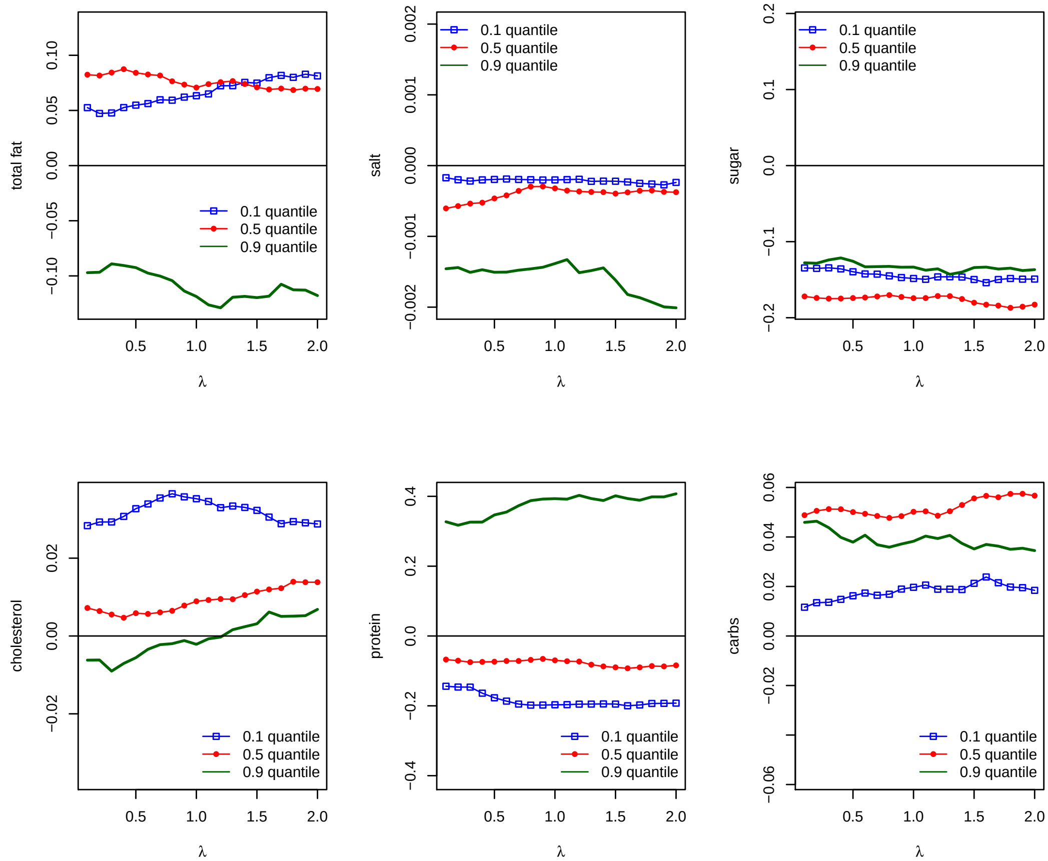

FIGURE 5.2. *Profile of preferences in terms of λ. The figure shows the difference between the estimated coefficients for female and male shoppers.*

We now turn to briefly document how preferences for nutrients vary by gender. Figure 5.2 reveals that these preferences differ substantially between males and females. We find that at the lower tail of the conditional expenditure distribution, females have stronger preferences for fat and cholesterol than males, while the opposite is true for salt and sugar. Furthermore, females have stronger preferences for protein at the upper quantiles of the distribution, while males prefer protein at the lower quantiles. This analysis reveals consumer preferences to be extremely heterogeneous and cautions against the use of demand models assuming homogeneous preferences.

specification used in (5.7). The null of exogeneity is not rejected in most of the cases with the exception of Income at the 0.5 and 0.9 quantiles when we compare pair of point estimates for $\lambda = \{1.5, 2\}$ with $\lambda = 0.01$. The results are available from the authors upon request.

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 an empirical application illustrating the practical implementation and use of the proposed methods.

In the 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.

Arellano, M., and O. Bover (1995): "Another look at the instrumental variable estimation of error-components models," *Journal of Econometrics*, 68(1), 29–51.

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

Belloni, A., D. Chen, V. Chernozhukov, and C. Hansen (2012): "Sparse Models and Methods for Optimal Instruments With an Application to Eminent Domain," *Econometrica*, 80(6), 2369–2429.

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., 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.

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., Fernández-Val, and A. Kowalski (2015): "Quantile Regression with Censoring and Endogeneity," *Journal of Econometrics, forthcoming*.

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 (2014): "Do Prices and Attributes Explain International Differences in Food Purchases?," *American Economic Review*, 104(3), 832–67.

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

Galvao, A. F., C. Lamarche, and L. R. Lima (2013): "Estimation of Censored Quantile Regression for Panel Data With Fixed Effects," *Journal of the American Statistical Association*, 108(503), 1075–1089.

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): "Penalized Quantile Regression with Semiparametric Correlated Effects: Applications with Heterogeneous Preferences," IZA Discussion Paper 7741.

Harding, M., and C. Lamarche (2014a): "Estimating and testing a quantile regression model with interactive effects," *Journal of Econometrics*, 178, Part 1(0), 101 – 113.

——— (2014b): "A Hausman-Taylor instrumental variable approach to the penalized estimation of quantile panel models," *Economics Letters*, 124(2), 176 – 179.

Harding, M., and M. Lovenheim (2014): "The Effect of Product and Nutrient-Specific Taxes on Shopping Behavior and Nutrition: Evidence from Scanner Data," working paper.

Hausman, J. A. (1978): "Specification Tests in Econometrics," *Econometrica*, 46(6), 1251–1271.

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

Hausman, J. A., and W. E. Taylor (1981): "Panel Data and Unobservable Individual Effects," *Econometrica*, 49 (6), 1377–1398.

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. (2014): *Analysis of Panel Data*. Cambridge University Press, New York, 3rd 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.

Levi, J., L. Segal, R. St. Laurent, and J. Rayburn (2014): "The State of Obesity," Discussion paper, Trust for America's Health.

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.

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. (2003): "Income Transfers and Assets of the Poor," *Review of Economics and Statistics*, 85(1), 63–76.

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,

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

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

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

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

□

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*

$$\boldsymbol{\Sigma}_0 = \lim_{\substack{T\to\infty \\ N\to\infty}} \frac{1}{TN} \begin{bmatrix} \Omega_{11}\boldsymbol{X}'\boldsymbol{W}_1'\boldsymbol{W}_1\boldsymbol{X} & \ldots & \Omega_{1J}\boldsymbol{X}'\boldsymbol{W}_1'\boldsymbol{W}_J\boldsymbol{X} \\ \vdots & \ddots & \vdots \\ \Omega_{1J}\boldsymbol{X}'\boldsymbol{W}_J'\boldsymbol{W}_1\boldsymbol{X} & \ldots & \Omega_{JJ}\boldsymbol{X}'\boldsymbol{W}_J'\boldsymbol{W}_J\boldsymbol{X} \end{bmatrix}$$

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

$$\boldsymbol{\Sigma}_2 = \lim_{\substack{T\to\infty \\ N\to\infty}} \frac{\Omega_m}{TN} \begin{bmatrix} \boldsymbol{X}'\boldsymbol{P}_1'\boldsymbol{P}_1\boldsymbol{X} & \ldots & \boldsymbol{X}'\boldsymbol{P}_1'\boldsymbol{P}_J\boldsymbol{X} \\ \vdots & \ddots & \vdots \\ \boldsymbol{X}'\boldsymbol{P}_J'\boldsymbol{P}_1\boldsymbol{X} & \ldots & \boldsymbol{X}'\boldsymbol{P}_J'\boldsymbol{P}_J\boldsymbol{X} \end{bmatrix}$$

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

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; $\boldsymbol{W}_j = \boldsymbol{I} - \boldsymbol{Z}\boldsymbol{P}_j$, $\boldsymbol{P}_j = (\boldsymbol{Z}'\boldsymbol{\Upsilon}_j\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{\Upsilon}_j$, $\boldsymbol{\Upsilon}_j = \boldsymbol{\Phi}_j(\boldsymbol{I} - \boldsymbol{\Lambda}_j)$, $\boldsymbol{\Phi}_j = diag(f_{it}(\xi_{it}(\tau_j)))$, $\boldsymbol{\Lambda}_j = diag(\boldsymbol{x}_i'(\boldsymbol{D}'\boldsymbol{Z}'\boldsymbol{\Phi}_j\boldsymbol{Z}\boldsymbol{D})^{-1}\boldsymbol{\Phi}_{ij}\boldsymbol{x}_i)$, $\boldsymbol{\Psi} = diag(f_a(0))$, and

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

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

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

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 $(\Delta_{0i},\Delta_1,\Delta_2)>0$, $\sup||k(\delta_{0i},\boldsymbol{\delta}_1,\boldsymbol{\delta}_2)-k(0,\mathbf{0},\mathbf{0})-\mathbb{E}(k(\delta_{0i},\boldsymbol{\delta}_1,\boldsymbol{\delta}_2)-k(0,\mathbf{0},\mathbf{0}))||=o_p(1)$, 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

$$
\begin{aligned}
&\mathbb{E}((k_t(\delta_{0i},\boldsymbol{\delta}_1,\boldsymbol{\delta}_2))-k_t(0,\mathbf{0},\mathbf{0})))= \\
&= \mathbb{E}\Bigg(-\frac{1}{\sqrt{N}}\sum_{j=1}^{J}\sum_{i=1}^{N}\omega_j\boldsymbol{x}_i\left(\psi_{\tau_j}\left(y_{it}-\frac{\delta_{0i}}{\sqrt{T}}-\boldsymbol{x}_{it}'\frac{\boldsymbol{\delta}_1(\tau_j)}{\sqrt{TN}}-\boldsymbol{x}_i'\frac{\boldsymbol{\delta}_2(\tau_j)}{\sqrt{N}}-\xi_{it}(\tau_j)\right)-\psi_{\tau_j}(y_{it}-\xi_{it}(\tau_j))\right)\Bigg) \\
&= -\frac{1}{\sqrt{N}}\sum_{j=1}^{J}\sum_{i=1}^{N}\omega_j\boldsymbol{x}_i\left(F_{it}\left(\xi_{it}(\tau_j)+\frac{\delta_{0i}}{\sqrt{T}}+\boldsymbol{x}_{it}'\frac{\boldsymbol{\delta}_1(\tau_j)}{\sqrt{TN}}+\boldsymbol{x}_i'\frac{\boldsymbol{\delta}_2(\tau_j)}{\sqrt{N}}\right)-\tau_j\right) \\
&= -\frac{1}{\sqrt{N}}\sum_{j=1}^{J}\sum_{i=1}^{N}\omega_j f_{it}(\xi_{it}(\tau_j))\boldsymbol{x}_i\left(\frac{\delta_{0i}}{\sqrt{T}}+\boldsymbol{x}_{it}'\frac{\boldsymbol{\delta}_1(\tau_j)}{\sqrt{TN}}+\boldsymbol{x}_i'\frac{\boldsymbol{\delta}_2(\tau_j)}{\sqrt{N}}\right)+o_p(1)
\end{aligned}
$$

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

$$\begin{aligned}
&\mathbb{E}((k(\delta_{0i}, \boldsymbol{\delta}_1)) - k(0, \mathbf{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) \\
&\quad + \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)) \Bigg(\frac{\delta_{0i}}{\sqrt{T}} + \boldsymbol{x}'_{it} \frac{\boldsymbol{\delta}_1(\tau_j)}{\sqrt{TN}} - \boldsymbol{x}'_i \boldsymbol{h}^{-1}_{jit} \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) \\
&\quad - \boldsymbol{x}'_i \boldsymbol{h}^{-1}_{jit} \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}} \Bigg) - 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}^{-1}_{jit} \omega_j \boldsymbol{x}_i \psi_{\tau_j} (y_{it} - \xi_{it}(\tau_j)) + \boldsymbol{x}'_i \frac{\boldsymbol{R}_{Nt}}{\sqrt{N}} \right)
\end{aligned}$$

where $\mu_{jit} = 1 - \boldsymbol{x}'_i \boldsymbol{h}^{-1}_{jit} \check{\boldsymbol{x}}_{ij}$, $f_i = T^{-1} \sum_{j=1}^{J} \sum_{t=1}^{T} \omega_j f_{it}(\xi_{it}(\tau_j))$, 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 $\mathbb{E}(k(\delta_{0i}, \boldsymbol{\delta}_1) - k(0, \mathbf{0})) = k(0, \mathbf{0})$. Therefore,

$$\begin{aligned}
&\frac{1}{\sqrt{T}} \sum_{j=1}^{J} \sum_{t=1}^{T} \omega_j f_{it} w_{itj} \frac{\delta_{0i}}{\sqrt{T}} + \frac{1}{\sqrt{T}} \sum_{j=1}^{J} \sum_{t=1}^{T} \omega_j f_{it} \mu_{jit} \boldsymbol{x}'_{it} \frac{\boldsymbol{\delta}_1(\tau_j)}{\sqrt{TN}} \\
&= \frac{1}{\sqrt{T}} \sum_{j=1}^{J} \sum_{t=1}^{T} \omega_j f_{it} \mu_{itj} \psi_{\tau_j} (y_{it} - \xi_{it}(\tau_j)) + 2 \frac{\lambda_T}{\sqrt{T}} \psi_{\tau_m}(a_i) - \frac{1}{\sqrt{T}} \sum_{j=1}^{J} \sum_{t=1}^{T} \omega_j f_{it} \boldsymbol{x}'_i \frac{\boldsymbol{R}_{Nt}}{\sqrt{N}} + \frac{\boldsymbol{R}_{Ti}}{\sqrt{T}}
\end{aligned}$$

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

$$\frac{\hat{\delta}_{0i}}{\sqrt{T}} = -f_{ji}^{-1}\left[\frac{1}{T}\sum_{j=1}^{J}\sum_{t=1}^{T}\omega_j f_{it}(\xi_{it}(\tau_j))\mu_{itj}\boldsymbol{x}_{it}'\right]\frac{\boldsymbol{\delta}_1(\tau_j)}{\sqrt{TN}} + f_{ji}^{-1}\Bigg(\frac{1}{T}\sum_{j=1}^{J}\sum_{t=1}^{T}\omega_j f_{it}\mu_{itj}\psi_{\tau_j}(y_{it}-\xi_{it}(\tau_j))$$
$$+2\frac{\lambda_T}{\sqrt{T}}\frac{\psi_{\tau_m}(a_i)}{\sqrt{T}} - \frac{1}{T}\sum_{j=1}^{J}\sum_{t=1}^{T}\omega_j f_{it}\boldsymbol{x}_i'\frac{\boldsymbol{R}_{Nt}}{\sqrt{N}}\Bigg) + \frac{\boldsymbol{R}_{Ti}}{\sqrt{T}} = -\sum_{j=1}^{J}\tilde{\boldsymbol{x}}_{it}(\tau_j)'\frac{\boldsymbol{\delta}_1(\tau_j)}{\sqrt{TN}} + \boldsymbol{r}_{it},$$

where $f_{ji} = T^{-1}\sum_j\sum_t \omega_j f_{it} w_{itj}$ and $\tilde{\boldsymbol{x}}_{it} = \sum_t \omega_j f_{it}\mu_{itj}\boldsymbol{x}_{it}'/Tf_i$. 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,

$$V_{TN}^{(1)}(\boldsymbol{\delta}_1) = -\sum_{j=1}^{J}\sum_{t=1}^{T}\sum_{i=1}^{N}\omega_j(\boldsymbol{x}_{it}' - \tilde{\boldsymbol{x}}_i(\tau_j)')(\boldsymbol{\delta}_1(\tau_j)/\sqrt{NT})\psi_{\tau_j}(y_{it}-\xi_{it}(\tau_j))$$
$$V_{TN}^{(2)}(\boldsymbol{\delta}_1) = \sum_{j=1}^{J}\sum_{t=1}^{T}\sum_{i=1}^{N}\omega_j\int_0^{v_{itj,TN}}(I(y_{it}-\xi_{it}(\tau_j)\leq s) - I(y_{it}-\xi_{it}(\tau_j)\leq 0))ds$$
$$V_{TN}^{(3)}(\boldsymbol{\delta}_1) = -\lambda_T\sum_{j=1}^{J}\sum_{i=1}^{N}\tilde{\boldsymbol{x}}_i(\tau_j)'(\boldsymbol{\delta}_1(\tau_j)/\sqrt{NT})\psi_{\tau_m}(a_i)$$
$$V_{TN}^{(4)}(\boldsymbol{\delta}_1) = \lambda_T\sum_{i=1}^{N}\int_0^{\tilde{\boldsymbol{x}}_i'\boldsymbol{\delta}_1/\sqrt{TN}}(I(a_i\leq s) - I(a_i\leq 0))ds$$

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 $(\Delta_{0i}, \Delta_1) > 0$,

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

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

$$\begin{aligned}
&\mathbb{E}((v(\eta_{0i},\boldsymbol{\eta}_1)) - v(0,\mathbf{0}))) = \\
&= \mathbb{E}\Bigg(-\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\psi_\tau\left(y_{it} - \frac{\eta_{0i}}{\sqrt{T}} - \boldsymbol{x}_{it}'\frac{\boldsymbol{\eta}_1(\tau)}{\sqrt{TN}} - \kappa_{it}(\tau)\right) + 2\frac{\lambda_T}{\sqrt{T}}\psi_\tau\left(\alpha_i + \frac{\eta_{0i}}{\sqrt{T}}\right) \\
&\quad + \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_\tau\left(y_{it} - \kappa_{it}(\tau)\right) - 2\frac{\lambda_T}{\sqrt{T}}\psi_\tau(\alpha_i)\Bigg) \\
&= -\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(F_{it}\left(\kappa_{it}(\tau) + \frac{\eta_{0i}}{\sqrt{T}} + \boldsymbol{x}_{it}'\frac{\boldsymbol{\eta}_1(\tau)}{\sqrt{TN}}\right) - \tau\right) + 2\frac{\lambda_T}{\sqrt{T}}\left(\tau - F_a\left(-\frac{\eta_{0i}}{\sqrt{T}}\right)\right) \\
&= -\frac{1}{\sqrt{T}}\sum_{t=1}^{T} f_{it}(\kappa_{it}(\tau))\left(\frac{\eta_{0i}}{\sqrt{T}} + \boldsymbol{x}_{it}'\frac{\boldsymbol{\eta}_1(\tau)}{\sqrt{TN}}\right) - \frac{\lambda_T}{\sqrt{T}}f_a(s_i)\frac{\eta_{0i}}{\sqrt{T}} + o(1)
\end{aligned}$$

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:

$$\begin{aligned}
V_{TN}^{(1)}(\boldsymbol{\eta}_1(\tau)) &= -\sum_{t=1}^{T}\sum_{i=1}^{N}(\boldsymbol{x}_{it}' - \dot{\boldsymbol{x}}_i')(\boldsymbol{\eta}_1(\tau)/\sqrt{NT})\psi_\tau(y_{it} - \kappa_{it}(\tau)) \\
V_{TN}^{(2)}(\boldsymbol{\eta}_1(\tau)) &= \sum_{t=1}^{T}\sum_{i=1}^{N}\int_0^{v_{it,TN}}(I(y_{it} - \kappa_{it}(\tau) \le s) - I(y_{it} - \kappa_{it}(\tau) \le 0))ds \\
V_{TN}^{(3)}(\boldsymbol{\eta}_1(\tau)) &= -\lambda_T\sum_{i=1}^{N}\dot{\boldsymbol{x}}_i'\left(\boldsymbol{\eta}_1(\tau)/\sqrt{NT}\right)\operatorname{sgn}(\alpha_i) \\
V_{TN}^{(4)}(\boldsymbol{\eta}_1(\tau)) &= \lambda_T\sum_{i=1}^{N}\int_0^{\dot{\boldsymbol{x}}_i'\frac{\boldsymbol{\eta}_1(\tau)}{\sqrt{TN}}}(I(\alpha_i \le s) - I(\alpha_i \le 0))ds
\end{aligned}$$

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 $\boldsymbol{H}_1 - \boldsymbol{\Sigma}_1 = \boldsymbol{L}'\boldsymbol{\Phi}\boldsymbol{L} - \boldsymbol{R}'\boldsymbol{\Upsilon}\boldsymbol{R} = \boldsymbol{L}'(\boldsymbol{\Phi} - \boldsymbol{\Upsilon})\boldsymbol{L} = \boldsymbol{L}'\boldsymbol{\Phi}\boldsymbol{\Lambda}\boldsymbol{L} = ||\boldsymbol{L}||^2_{\boldsymbol{\Phi}\boldsymbol{\Lambda}} > \boldsymbol{0}$, 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 $\mathbb{R}_+$. □

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

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

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,

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

□

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,

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

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, $\text{trAVar}(\hat{\boldsymbol{\beta}}(\tau,\lambda^*)) < \text{trAVar}(\hat{\boldsymbol{\beta}}(\tau,\lambda))$. □