Group Number Selection for Two-Dimensional Grouping in Panel Quantile Regression

Bachelor Thesis for the BSc in Econometrics and Economics

Getoar Sopa

512673

Abstract

This thesis introduces two novel procedures to estimate the number of groups in a two-dimensionally grouped quantile regression setting with separately grouped slopes and fixed effects, as developed in Leng et al. (2021). The first of these procedures involves selecting the group numbers according to a Weighted Information Criterion, whereas the second of these procedures involves selecting the group numbers according to an LM testing procedure. I compare the performance of these two procedures to two pre-existing procedures: the composite IC procedure, which equally aggregates information across all quantiles into one IC value, and the Max-IC procedure, which makes quantile-wise group number calculations and chooses the maximally estimated group numbers across all quantiles. I find that the Weighted Information Criterion is unstable and yields imprecise results, whereas the testing procedure lacks power under small sample sizes or weak group separation. The Max-IC procedure is found to generally be the safest choice, though it lacks accuracy compared to the composite IC procedure. Finally, I apply all four methods to data on the productivity of US states and find that all IC procedures agree on the number of groups, but that the testing procedure yields very different estimates. The results of this final analysis suggest that the elasticity of production with respect to public capital is heterogeneous across states.

Supervisor: dr. Wendun Wang

Second Assessor: dr. Carlo Cavicchia

Erasmus School of Economics,

Erasmus University Rotterdam

July 1, 2022

Contents

1 Introduction

Since the introduction of the basic quantile regression model by Koenker and Bassett Jr (1978), increasingly sophisticated and specialized models for estimating quantile models have appeared in the literature. Some of these additions involve the clustering of units into distinct groups, which are then used to construct regression parameters. One such recent model is a panel quantile regression model with grouped cross-sectional fixed effects, introduced by Gu and Volgushev (2019). This paper also introduced a model in which both cross-sectional fixed effects and the slope coefficients enjoy group heterogeneity, but they assume one-dimensional grouping, i.e. units that share a fixed effect also share a slope coefficient.

More recently, Leng et al. (2021) introduced a panel quantile regression model with two-dimensional group clustering for cross-sectional fixed effects and slope coefficients, respectively, with the additional inclusion of time fixed effects. This versatile model estimates all regression coefficients as well as the two-dimensional group membership parameters in an iterative manner. This model is not fully self-sufficient, however, as it requires the total number of cross-sectional fixed effect and slope groups in the data to be known. In practice, these must be estimated from the relevant data. If these group numbers are set below their true levels in the quantile regression, the coefficient estimation ceases to be consistent and unbiased. This is because units belonging to separate slope groups will be incorrectly grouped together, resulting in underfitting of the model (Zhang et al., 2019). On the other hand, setting these numbers (far) above their true levels when estimating the model yields inefficient estimates. However, they remain consistent, and so in practice selecting the group numbers conservatively is less harmful to the coefficient estimates, given that the sample is large enough. Choosing the number of groups in this model is therefore an important balancing act for finite-sample studies.

In their paper, Leng et al. (2021) propose two strategies to estimate these group numbers. The primary strategy they propose is a composite Information Criterion approach, which aggregates information across all quantiles to select optimal group numbers and yields consistent estimates. This is the most common approach in the grouped quantile regression literature, used, for example, in Bonhomme and Manresa (2015), Su et al. (2016), and Liu et al. (2020). Several variants of such a composite IC are possible. Some authors, such as Lee et al. (2014), use a BIC-like IC which takes the natural logarithm of the objective value. This is also the approach

I take in this thesis.

To prevent quantiles with weak group separation from contaminating the IC, resulting in the underestimation of the group numbers, Leng et al. (2021) suggest as a secondary procedure computing the IC value for each quantile separately, and then selecting the maximum estimated group numbers across all quantiles. This procedure never selects group numbers smaller than the composite approach but is much likelier to overestimate the group numbers, and may therefore be favored in settings with strongly varying degrees of group separation across quantiles, or when efficiency is a secondary concern. In the remainder of this thesis, I refer to this method of selecting the number of groups as the Max-IC procedure.

In sum, there are two procedures established in the literature to determine the number of groups in the model of Leng et al. (2021). Both of these procedures use an Information Criterion, the tuning parameter of which is sensitive to precise model specification. Moreover, one of these procedures is prone to underestimation of the number of groups if some quantiles suffer weak group separation, while the other is prone to regular overestimation of the number of groups. In addition, the accuracy of these two procedures in determining the correct number of groups has not yet been compared - only the accuracy of the composite IC procedure has been investigated. This motivates the three main contributions of this thesis.

The first of these contributions is the development of an Information Criterion-based procedure designed to not allow quantiles with weak group separation to contaminate the Information Criterion as much as in the composite IC while moderating the tendency to overestimate the group numbers. The Information Criterion used in this procedure, which I refer to as the Weighted Information Criterion (WIC), is a simple adaptation of the composite IC. Whereas in the composite IC approach the contributions of all quantiles are aggregated equally, the WIC assigns weights to the contribution of quantiles to the IC based on the degree of group separation present in the quantiles. This should, if the weights are chosen appropriately, reduce the threat of IC contamination while still allowing for an aggregated IC function.

The second main contribution of this thesis is the introduction of a procedure that does not make use of an Information Criterion at all. This allows one to determine the group numbers in a more general quantile regression setting with two-dimensional grouping, without requiring the adjustment of the information criteria parameters to accommodate slight changes to model specification. The IC-less method I employ to determine the number of groups in our setting is a testing-based procedure. Earlier papers in the literature, such as Lu and Su (2017), have touched upon the use of testing to determine group numbers, but none have been used in a quantile regression or two-dimensional clustering setting. I adapt the procedure proposed in Lu and Su (2017) to our setting, which involves a sequence of group number-specific LM-type tests using residuals from the median regression. I additionally lay the foundation for a testing procedure that incorporates information from the residuals across all quantiles, but leave the derivation of a test statistic for that case to future work.

The final novel contribution of this thesis is the comparison of the accuracy of the group number selection procedures discussed in the previous paragraphs in several simulation studies. Moreover, I investigate in what settings each of the methods may be preferred over the others, and I assess the effect that over- or underestimating the number of latent groups has on slope coefficient estimates. I find that the Weighted Information Criterion procedure yields unstable results and is very sensitive to the exact model specification and that as a result, the Max-IC procedure is generally the preferred and safer option. This latter procedure is found to work as intended, with consistent overestimation of the group numbers, and only rare underestimation. Therefore, in settings with complex data, the use of the Max-IC procedure may be preferred over the composite IC as well. Finally, the LM-based testing procedure is found to select the number of slope groups accurately in the presence of enough slope group separation or with a very large sample size. However, it consistently underestimates the number of fixed effect groups in a finite-sample setting due to the relatively small impact underspecification of the number of fixed effect groups has on the slope coefficients. In addition, it is found to be very susceptible to underestimation of the number of slope groups in the presence of weak slope group separation.

Finally, to illustrate the findings of this thesis, I apply all four group number estimation methods to a quantile regression analysis of productivity data from 48 US states across a 17-year period to examine the elasticity of state productivity with respect to public capital. I find that (i) not all group number estimation procedure agree on the estimated group numbers and (ii) there are indications of heterogeneous public capital elasticities. This result suggests that the choice of group number estimation procedure can have significant consequences in practical inference, and that future research into the elasticity of aggregate production with respect to public capital must take possibly heterogeneous effects into account.

2 Literature Review

This research is generally related to work on linear panel models with some form of grouped coefficient heterogeneity, such as Lin and Ng (2012), which considers heterogeneous slope, but not fixed effect, coefficients, and approaches this heterogeneity issue with both a K-means clustering approach as well as a panel threshold approach, and Bonhomme and Manresa (2015), which separately considers both the case of grouped fixed effects as well as grouped slope coefficients, again using a K-means-like approach for model estimation. Although these models focus on mean-based rather than quantile regression, and they do not consider two-dimensional grouping in both the fixed effects as well as the slope coefficients simultaneously, the problems they address are not fundamentally different than the problems addressed in this thesis, and many of their findings can be translated to this setting. More specifically related is, of course, the literature on panel quantile structure models, culminating in the two-dimensionally grouped panel quantile model with time fixed effects proposed in Leng et al. (2021), which this thesis uses as a framework. Perhaps unsurprisingly, many of the results found in the quantile panel case have analogues in the linear panel regression case.

In addition, the existing literature on selecting the number of groups in any panel regression setting with heterogeneous coefficients is directly related to the work in this thesis. A large part of this literature is devoted to the use of Information Criteria; examples are Gu and Volgushev (2019), Su et al. (2016), and Su and Ju (2018). Naturally, there are also many group number estimation methods that do not involve the use of Information Criteria. An example mentioned in the introduction is a testing-based procedure to determine the number of slope groups in a panel regression setting, developed in Lu and Su (2017). Other examples are discussed in Yu et al. (2022), which assesses several methods to determine the number of groups in a panel quantile setting with either grouped slope coefficients or grouped fixed effects, but not both. In their paper, the authors consider for the selection of the number of slope groups a cross-validation procedure that exploits stability of group assignment under the true number of groups, adopted in Zhang et al. (2019) from Wang (2010), as well as a modified version of the aforementioned method in which spectral clustering is used, and a heuristic that chooses the number of groups by maximizing the relative eigengap of a modified graph Laplacian. For a detailed discussion on spectral clustering and graph Laplacians, I refer the reader to Von Luxburg (2007). For the case of homogeneous slope coefficients but grouped fixed effects, the authors also consider the Information Criterion proposed in Gu and Volgushev (2019). In their simulation studies, they find that no single method outperformed the other considered methods in all settings. Although this thesis is closely related to Yu et al. (2022), I consider the more general model with separately grouped fixed effects and slope coefficients, and assess, for the most part, different group number selection procedures, two of which are first introduced in this thesis.

Finally, the finite sample simulation-based assessment of the model's performance with under- or overspecified group numbers considered in this thesis is related to the theoretical analyses performed in Angrist et al. (2006), which examine the theoretical consequences of omitted variables in a general quantile regression setting (noting that with group number underspecification, at least one variable is missing from the model), as well as to the theoretical analysis of Liu et al. (2020), which considers overestimation of the number of groups. Both of these papers, either directly or indirectly, consider the theoretical consequences of over- or underestimation of the number of groups in a panel setting with grouped coefficients as N and T pass to infinity, while in this thesis I examine the practical consequences of such incorrect group number estimations.

3 Methodology

3.1 Model Set-up

The setting I consider in this thesis is the same as in Leng et al. (2021). In this panel setting, there are N units and T periods, and units are partitioned into two distinct sets of groups. First, there are H cross-section fixed effects groups, and all units in the same fixed effect group share a common intercept. Second, there are G slope coefficient groups, and all units in the same slope coefficient group share common slopes.

In this setting, and given a dataset, let y_{it} denote the one-dimensional dependent variable and x_{it} denote the p exogenous regressors of unit i at time t. In this model, there is a three-sided estimation interest. First, one is interested in estimating which of the H cross-sectional fixed effect groups each unit i belongs to, denoted by h_i. Similarly, one is interested in which of the G slope coefficient groups each unit i belongs to, denoted by g_i. Finally, given these group membership variables, one is interested in the actual quantile regression coefficients:

$$Q_\tau(y_{it}|x_{it}) = \alpha_{h_i}(\tau) + \lambda_t(\tau) + x_{it}^{'}\beta_{g_i}(\tau), \quad i = 1, ..., N, \quad t = 1, ..., T. \tag{1}$$

Here, $Q_\tau(y_{it}|x_{it})$ denotes the conditional τ'th quantile of the dependent variable y_{it} given the p exogenous regressors x_{it}, and $\alpha_{h_i}(\tau)$ represents the cross-section fixed effects of the units belonging to group h_i, $\lambda_t(\tau)$ represents the time fixed effects, and $\beta_{g_i}(\tau)$ denotes the $p \times 1$ vector of slope coefficients of units belonging to group g_i, all at the τ'th quantile.

To estimate this model, first let γ_g denote the vector of all N slope group membership parameters, i.e., $\gamma_g = \{g_1, ..., g_N\}$ and γ_h denote the vector of all N cross-sectional fixed effect group membership parameters, i.e. $\gamma_h = \{h_1, ..., h_N\}$. Finally, let $\theta(\tau_k)$ denote the vector $(\beta_1(\tau_k)^{'}, ..., \beta_G(\tau_k)^{'}, \alpha_1(\tau_k), ..., \alpha_H(\tau_k),\ \lambda_1(\tau_k), ..., \lambda_T(\tau_k))$, and $\boldsymbol{\theta(\tau)} = (\theta(\tau_1), ..., \theta(\tau_K))$. In this thesis, the quantiles $\tau_k = \frac{k}{K+1}$ with $1 \leq k \leq K = 9$ are maintained throughout all analyses, but any number and form of valid quantiles is applicable. Joint estimation of γ_g, γ_h, and the coefficients of model (1) can then be performed using Algorithm 1 proposed in Leng et al. (2021), which can also be found below. Step 1 of this algorithm is estimated with the usual quantile regression estimator discussed in Koenker and Hallock (2001), with the inclusion of the relevant variables.

Note that initial estimates for γ_g and γ_h are necessary in the initialization of Algorithm 1. These may come from other models or algorithms, but in this thesis I choose to run Algorithm 1 100 times with new random initial group membership parameters each run. Then, for each of the resulting set of estimates, I calculate the associated objective function

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\rho_{\tau_k}(y_{it} - \alpha_{h_i}(\tau_k) - \lambda_t(\tau_k) - x_{it}^{'}\beta_{g_i}(\tau_k)),$$

where $\rho_\tau(u) = [\tau - I(u < 0)]u$ denotes the check function, and I choose the estimated parameters associated with the lowest objective value. This entire process therefore yields one set of parameter estimates.

When estimating this model, the number of cross-sectional fixed effect groups H and the number of slope groups G are usually unknown, and highly dependent on the particular data used. Thus, they must be estimated from the data at hand before the coefficients in equation (1) can be estimated by use of Algorithm 1. As discussed in the Introduction, underestimation of these two values leads to inconsistent coefficient estimates, whereas overestimation of these values leads to inefficient coefficient estimates.

Algorithm 1 of Leng et al. (2021) for the estimation of equation (1)

Set $s = 0$ and let $\gamma_g^{(0)}$ and $\gamma_h^{(0)}$ denote the initial group membership parameters for the g-group and h-group respectively.

repeat

1. Given group membership estimates of the previous step $\gamma_g^{(s)}$ and $\gamma_h^{(s)}$, estimate the regression quantile parameters as:

$$\boldsymbol{\theta^{(s)}(\tau_k)} = \underset{\boldsymbol{\theta(\tau_k)}}{\text{argmin}} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \rho_{\tau_k}(y_{it} - \alpha_{h_i^s}(\tau_k) - \lambda_t(\tau_k) - x_{it}^{'}\beta_{g_i^s}(\tau_k))$$

for $k = 1, ..., K$.

2. Given regression quantile parameters $\boldsymbol{\theta^s(\tau)}$ and h-group membership parameter estimates $\gamma_h^{(s)}$, estimate g-group membership parameters as:

$$g_i^{(s+1)} = \underset{g \in \{1,...,G\}}{\text{argmin}} \frac{1}{T} \sum_{t=1}^{T} \sum_{k=1}^{K} \rho_{\tau_k}(y_{it} - \alpha_{h_i^s}(\tau_k) - \lambda_t(\tau_k) - x_{it}^{'}\beta_g^{(s)}(\tau_k))$$

for $i = 1, ..., N$.

3. Given regression quantile parameters $\boldsymbol{\theta^s(\tau)}$ and g-group membership parameter estimates $\gamma_h^{(s+1)}$, estimate h-group membership parameters as:

$$h_i^{(s+1)} = \underset{h \in \{1,...,H\}}{\text{argmin}} \frac{1}{T} \sum_{t=1}^{T} \sum_{k=1}^{K} \rho_{\tau_k}(y_{it} - \alpha_h(\tau_k) - \lambda_t(\tau_k) - x_{it}^{'}\beta^{(s)}_{g_i^{(s+1)}}(\tau_k))$$

for $i = 1, ..., N$.

4. Set $s = s + 1$.

until The quantile regression estimates $\boldsymbol{\theta^s(\tau)}$ have converged numerically.

To limit the size of this estimation process while excluding the possibility that there is guaranteed underestimation, G and H are usually chosen from the range $1 \leq G \leq G_{max}$ and $1 \leq H \leq H_{max}$, with G_{max} and H_{max} chosen conservatively large. In this thesis, two existing and two novel methods of choosing G and H in the context of two-dimensionally grouped quantile regression are used. The pre-existing composite IC and Max-IC procedures are discussed in section 3.2.1, whereas the weighted IC and testing procedures, first introduced in this thesis, are discussed in sections 3.2.2 and 3.3.

3.2 IC-based Group Number Selection

3.2.1 Pre-Existing IC Procedures

Besides proposing an estimation strategy for the quantile regression presented in equation (1), Leng et al. (2021) also propose an IC-based strategy to estimate the two group sizes, G and H. This procedure evaluates an Information Criterion for every choice of (G, H) with $1 \leq G \leq G_{max}$ and $1 \leq H \leq H_{max}$, and selects the (G, H) associated with the minimal IC value. The Information Criterion utilized in this thesis is as follows:

$$IC(G,H) = log\left[\frac{1}{NT}\sum_{k=1}^{K}\sum_{i=1}^{N}\sum_{t=1}^{T}\rho_{\tau_k}\left(y_{it} - \hat{\alpha}_{\hat{h}_i}^{(G,H)}(\tau_k) - \hat{\lambda}_t^{(G,H)}(\tau_k) - x_{it}^{'}\hat{\beta}_{\hat{g}_i}^{(G,H)}(\tau_k)\right)\right] + \kappa K(pG+H), \tag{2}$$

where κ is a tuning parameter and the superscript (G, H) denotes estimates obtained when using the group number pair (G, H). I maintain

$$\kappa = 0.5 \cdot log(NT)/NT$$

in this thesis, as that value has yielded favorable results in experiments. This parameter is quite sensitive to changes in the data generating process, and other tuning parameters may generally yield more accurate results.

Note that in IC (2) a composite approach is chosen which sums over results across all considered quantiles. Although this is generally favorable, as it yields more information for the group number estimation, Leng et al. (2021) indicate that if there is weak group separation at several quantiles, then the IC contribution of those quantiles may adversely affect the performance of the composite IC. This is because weak group separation at a few quantiles, but not at the others, has as a consequence that increasing the group numbers to the true levels does not improve the contribution of those quantiles to the objective function much. This may result in the right-hand side of (2) growing faster than the left-hand side shrinks as the numbers of groups increase to their true levels, and, as a consequence, lead to underestimation of the group numbers.

To remedy this, the authors suggest an additional but related method to determine the level of G and H: the Max-IC procedure. In this approach, (2) is evaluated separately at each quantile, yielding K sets of optimal group numbers. (G, H) will then be selected as the maximum across the respective sets of K estimated group numbers. Concretely, for each $k \in \{1, .., K\}$, one may calculate, for each pair of (G, H), with $1 \leq G \leq G_{max}$ and $1 \leq H \leq H_{max}$, the following quantile-specific IC:

$$IC_k(G,H) = log\left[\frac{K}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\rho_{\tau_k}\left(y_{it} - \hat{\alpha}_{\hat{h}_i}^{(G,H)}(\tau_k) - \hat{\lambda}_t^{(G,H)}(\tau_k) - x_{it}^{'}\hat{\beta}_{\hat{g}_i}^{(G,H)}(\tau_k)\right)\right] + \kappa K(pG+H). \quad (3)$$

Note that in this quantile-specific IC function, each quantile objective value is multiplied by a factor K. This allows the use of a similar tuning parameter as for the composite IC. For comparison's sake, I elect to maintain the same tuning parameter as before - small adjustments might, in practice, lead to better results.

Once all combinations of $IC_k(G, H)$ are calculated, there is one pair (G_k, H_k) associated with the minimal IC_k value for each k. This results in the K estimated group number pairs (G_1, H_1) up to (G_K, H_K), and G and H are then chosen as $G = max\{G_1, ..., G_K\}$ and $H = max\{H_1, ..., H_K\}$. This method eliminates the risk of IC contamination by quantiles with weak group separation and is, in general, much less susceptible to group number underestimation than its composite counterpart, but it also yields less efficient regression estimation than the composite approach.

3.2.2 The Weighted IC Procedure

An alternative way to deal with the issue of IC contamination by weakly separated quantiles is by weighting quantiles according to their degree of group separation - an approach not yet considered in the existing literature. When weighting quantiles' contribution to the IC, multiple possibilities for such a weighting procedure arise. It is, for instance, possible to either weight the residuals directly or to weight the log contribution of each quantile. As weighting the residuals directly would involve a non-linearity due to the logarithm, I choose in this thesis to weight the log-contribution of the quantiles. Generally speaking, this leads to the following Weighted Information Criterion (WIC):

$$WIC(G,H) = \sum_{k=1}^{K} w_k\left(log\left[\frac{K}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\rho_{\tau_k}\left(y_{it} - \hat{\alpha}_{\hat{h}_i}^{(G,H)}(\tau_k) - \hat{\lambda}_t^{(G,H)}(\tau_k) - x_{it}^{'}\hat{\beta}_{\hat{g}_i}^{(G,H)}(\tau_k)\right)\right]\right) + \kappa K(pG+H), \quad (4)$$

where w_k denotes the weighting term. To maintain consistency across our methods, I once again elect to use the same tuning parameter as before, and to multiply the quantile objective function by a factor K. Under the assumption that $\sum_{k=1}^{K} w_k = 1$, the WIC should then perform similarly to the IC and Max-IC procedures. Of course, if weights are not chosen in this way, the inflation factor in the quantile objective function may also be adjusted or removed entirely, along with an appropriate accomodating change in the tuning parameter.

There are many ways to select weighting functions that both satisfy the aforementioned constraint and weight quantiles with large group separation more heavily. One such weighting function, and the one I maintain throughout the rest of this thesis, is the following:

$$w_k = \frac{\frac{1}{G}\sum_{g=1}^{G}||\hat{\beta}_g^{(G,H)}(\tau_k) - \hat{\beta}_{mean}^{(G,H)}(\tau_k)|| + \frac{1}{H}\sum_{h=1}^{H}||\hat{\alpha}_h^{(G,H)}(\tau_k) - \hat{\alpha}_{mean}^{(G,H)}(\tau_k)||}{\sum_{k=1}^{K}(\frac{1}{G}\sum_{g=1}^{G}||\hat{\beta}_g^{(G,H)}(\tau_k) - \hat{\beta}_{mean}^{(G,H)}(\tau_k)|| + \frac{1}{H}\sum_{h=1}^{H}||\hat{\alpha}_h^{(G,H)}(\tau_k) - \hat{\alpha}_{mean}^{(G,H)}(\tau_k)||)}. \quad (5)$$

The result of choosing such weights is that those quantiles with strong group separation dominate the left-hand term of the Information Criterion, which should lead to combinations with higher group numbers being favored. At the same time, by incorporating information from all quantiles simultaneously, the efficiency loss should not be as large as with the Max-IC procedure. Note that the WIC with these weights can not be used to test (G, H) = (1, 1) against other alternatives. One may either compare the standard composite IC of the (1, 1) case against the WIC of the alternatives or omit the (1, 1) case entirely from the WIC comparison.

In practice, quantile regressions are often run using a default constant term. In this case, the effect of one cross-sectional fixed effect group and one time fixed effect are not separately identified; I assume that these are α_1 and λ_1. In order not to allow the time fixed effect to influence the weighting procedure, in this setting the weights should be slightly adjusted as follows, if $H > 1$:

$$w_k = \frac{\frac{1}{G}\sum_{g=1}^{G}||\hat{\beta}_g^{(G,H)}(\tau_k) - \hat{\beta}_{mean}^{(G,H)}(\tau_k)|| + \frac{1}{H-1}\sum_{h=2}^{H}||\hat{\alpha}_h^{(G,H)}(\tau_k) - \hat{\alpha}_{mean}^{(G,H)}(\tau_k)||}{\sum_{k=1}^{K}(\frac{1}{G}\sum_{g=1}^{G}||\hat{\beta}_g^{(G,H)}(\tau_k) - \hat{\beta}_{mean}^{(G,H)}(\tau_k)|| + \frac{1}{H-1}\sum_{h=2}^{H}||\hat{\alpha}_h^{(G,H)}(\tau_k) - \hat{\alpha}_{mean}^{(G,H)}(\tau_k)||)}, \quad (6)$$

or, if $H = 1$, and $G > 1$:

$$w_k = \frac{\sum_{g=1}^{G}||\hat{\beta}_g^{(G,H)}(\tau_k) - \hat{\beta}_{mean}^{(G,H)}(\tau_k)||}{\sum_{k=1}^{K}\sum_{g=1}^{G}||\hat{\beta}_g^{(G,H)}(\tau_k) - \hat{\beta}_{mean}^{(G,H)}(\tau_k)||}. \quad (7)$$

3.3 Testing-based Group Number Selection

The methods for determining the number of groups in this two-dimensional case outlined in the previous subsections have in common that they rely on an Information Criterion. Although these procedures yield reliable and consistent results if properly tuned, it is precisely this requirement of tuning the penalty term that complicates their use in practice. An entirely different method of selecting the number of groups is by means of a testing procedure, which can be universally applied without adjustments. For our model, I employ an adapted version of the sequential LM-based testing procedure first proposed in Lu and Su (2017), which is designed for settings with a one-dimensional group structure in a mean-based regression model.

This setup involves a sequence of LM tests, the statistic of which is worked out in the following paragraphs. Given this test statistic, the sequence of tests is performed as follows. First, set $H = 1$ and $G = 1$, estimate the associated quantile regression model, and test the hypothesis $H_{null} : G = 1,\ H = 1$ versus $H_{alternative} : 1 < G \leq G_{max}$ or $1 < H \leq H_{max}$ or both. If the test rejects this null hypothesis, exhaust the group number combinations for $G + H = 3$ and test the associated null hypotheses. Continue this entire procedure until the null hypothesis is no longer rejected, at which point the procedure's estimated G and H follow from the null hypothesis. This therefore involves a sequence of tests, where one sequentially exhausts all possibilities for $G + H = 2$, then $G + H = 3$, then $G + H = 4$, and so forth, up until $G + H = G_{max} + H_{max}$, or until the associated null hypothesis is no longer rejected, whichever occurs sooner. If the null is not rejected for any combination of G and H, one may either accept $(G, H) = (G_{max}, H_{max})$, or increase G_{max} and H_{max}, and continue the testing procedure for higher values of G and H. It may occur that for $G_1 \neq G_2$, $H_1 \neq H_2$, but $G_1 + H_1 = G_2 + H_2$, the null hypothesis is rejected for neither (G_1, H_1) nor (G_2, H_2). In this case, one may select the combination of G and H that yields the test statistic that is smallest in magnitude.

I now derive the test used for any given combination (G, H). First, given (G, H), estimate $\hat{\alpha}^{(G,H)}_{\hat{h}_i}(\tau_k)$, $\hat{\lambda}^{(G,H)}_t(\tau_k)$, *and* $\hat{\beta}^{(G,H)}_{\hat{g}_i}(\tau_k)$ for all values of k, *as well as* for the value $\tau = 0.5$, using Algorithm 1 of Leng et al. (2021). Note that in the following, I suppress (G, H) in the notation of the coefficient estimates.

The functional form of y_{it} in our Quantile Regression Model framework, where I assume a location-scale shift model, is as follows:

$$y_{it} = \alpha_{h_i} + \lambda_t + x_{it}\beta_{g_i} + (1 + \psi x_{it})\varepsilon_{it}. \tag{8}$$

I assume no particular distribution for ε_{it}, but ε_{it} must satisfy both $E[\varepsilon_{it}|X_{it}] = 0$ as well as $Q_{0.5}[\varepsilon_{it}|X_{it}] = 0$, where X_{it} denotes all regressors. In addition, given the coefficient estimates resulting from Algorithm 1, one can calculate the model residuals as follows:

$$\hat{\varepsilon}_{it}(\tau) = y_{it} - \hat{\alpha}_{\hat{h}_i}(\tau) - \hat{\lambda}_t(\tau) - x_{it}\hat{\beta}_{\hat{g}_i}(\tau). \tag{9}$$

Equations (8) and (9), when combined, straightforwardly yield the following expression for $\hat{\varepsilon}_{it}(\tau)$:

$$\hat{\varepsilon}_{it}(\tau) = \varepsilon_{it} + (\alpha_{h_i} - \hat{\alpha}_{\hat{h}_i}(\tau)) + (\lambda_t - \hat{\lambda}_t(\tau)) + (\beta_{g_i} + \psi\varepsilon_{it} - \hat{\beta}_{\hat{g}_i}(\tau))x_{it}. \tag{10}$$

As such, following the consistency of the two-dimensionally grouped quantile regression estimator, and given that $G \geq G_{true}$ and $H \geq H_{true}$,

$$\hat{\varepsilon}_{it}(\tau) \to \varepsilon_{it} + (\beta_{g_i} + \psi\varepsilon_{it} - \hat{\beta}_{\hat{g}_i}(\tau))x_{it} \text{ for } N, T \to \infty.$$

A consequence of the assumption that $E[\varepsilon_{it}|X_{it}] = 0$, is that $E[\hat{\varepsilon}_{it}(\tau)|X_{it}] \to (\beta_{g_i} - \hat{\beta}_{\hat{g}_i}(\tau))x_{it}$ for $N, T \to \infty$. Therefore, $E[\hat{\varepsilon}_{it}(\tau)|X_{it}] \to 0$ if and only if $\hat{\beta}_{\hat{g}_i}(\tau) \to \beta_{g_i}$. Given that $Q_{0.5}[\varepsilon_{it}|X_{it}] = 0$ and that

$$\hat{\beta}_g(\tau) \to \beta_g + \psi Q_\tau[\varepsilon_{it}] \text{ and } \hat{g}_i \to g_i \quad \text{if } G \geq G_{true},\ H \geq H_{true,},\ N, T \to \infty$$

it follows that $E[\hat{\varepsilon}_{it}(\tau)|X_{it}] \to 0$ for $\tau = 0.5$ if $G \geq G_{true}$ and $H \geq H_{true}$. These findings imply that one can test $G \geq G_{true}$ and $H \geq H_{true}$ by running the following auxiliary regression:

$$\hat{\varepsilon}_{it}(\tau = 0.5) = \nu_i + \phi_i^{'}(\tau = 0.5)x_{it} + \eta_{it}(\tau = 0.5), \tag{11}$$

and testing the null hypothesis $\mathbb{H}_0 : \phi_i^{'}(\tau = 0.5) = 0 \quad \forall i = 1, ..., N$. In the following paragraphs, I suppress the notation $\tau = 0.5$.

The hypothesis formulated in the preceding paragraph can be tested using the LM test statistic developed in Lu and Su (2017). To derive this test statistic, let $M_0 = I_T - (1/T)\boldsymbol{i}_T\boldsymbol{i}_T^{'}$, where I_T denotes the $T \times T$ identity matrix and $\boldsymbol{i}_T$ denotes the $T \times 1$ vector of ones. Moreover, let $H_i = M_0 x_i (x_i^{'} M_0 x_i)^{-1} x_i^{'} M_0$, where x_i denotes the $T \times p$ vector $(x_{i1}, x_{i2}, ..., x_{iT})^{'}$, and let $h_{i,ts}$ denote the element in the $t'th$ row and $s'th$ column of H_i. Finally, let $\hat{\Omega} = (1/T)x_i^{'} M_0 x_i$, and $\hat{b}_i = M_0 x_i \hat{\Omega}_i^{-1/2}$.

The unscaled LM statistic can then be calculated as

$$LM = \sum_{i=1}^{N} \hat{\varepsilon}_i^{'}(\tau) M_0 x_i (x_i^{'} M_0 x_i)^{-1} x_i M_0 \hat{\varepsilon}_i(\tau),$$

where $\hat{\varepsilon}_i(\tau)$ denotes the $T \times 1$ vector $(\hat{\varepsilon}_{i1}(\tau), ..., \hat{\varepsilon}_{iT}(\tau))$. This statistic must be scaled before it follows a particular distribution under H_0. For this purpose, define

$$\hat{B} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{\varepsilon}_{it}^2(\tau) h_{i,tt},$$

and

$$\hat{V} = \frac{4}{T^2 N} \cdot \sum_{i=1}^{N} \sum_{t=2}^{T} \left(\hat{\varepsilon}_{it}(\tau) \hat{b}_{it}^{'} \sum_{s=1}^{t-1} \hat{b}_{is} \hat{\varepsilon}_{is}(\tau) \right)^2,$$

where $\hat{b}_{ir}$ denotes the $r'th$ row of $\hat{b}_i$. Then,

$$(\frac{1}{\sqrt{N}} LM - \hat{B}) / \sqrt{\hat{V}} \to N(0,1) \text{ as } N, T \to \infty,$$

where this asymptotic distribution holds under certain technical assumptions discussed extensively in Lu and Su (2017).

Note that, in principle, this procedure works for any value of τ for which $Q_\tau[\varepsilon_{it}|X_{it}] = 0$, with the sole change to the testing procedure, in this case, being that the residuals resulting from the τ-level regression must be used instead of the residuals resulting from the median regression. Of course, most common distributions do not allow for multiple quantiles to equal each other, so in practice, this procedure can be performed for only one series of residuals per DGP.

To remedy the aforementioned limitation, alternative tests that incorporate information from all available series of residuals *are* possible, and they depend on the properties of quantile regression estimation. To derive these tests, first note that, for any value of $\tau \in (0,1)$,

$$Q_\tau(\hat{\varepsilon}_{it}(\tau)) = Q_\tau(\varepsilon_{it}) + (\alpha_{h_i} - \hat{\alpha}_{\hat{h}_i}(\tau)) + (\lambda_t - \hat{\lambda}_t(\tau)) + (\beta_{g_i} + \psi Q_\tau(\varepsilon_{it}) - \hat{\beta}_{\hat{g}_i})x_{it}.$$

If $G \geq G_{true}$ and $H \geq H_{true}$, the consistency of the estimates resulting from Algorithm 1 yields:

$$Q_\tau(\hat{\varepsilon}_{it}(\tau)) \rightarrow Q_\tau(\varepsilon_{it}) \quad \forall \tau \in (0,1).$$

This, therefore, motivates testing $G \geq G_{true}$ and $H \geq H_{true}$ by testing $H_0 : \phi_i^{'} = 0 \; \forall i = 1, ..., N,$ and for all values of τ, in the following quantile regression:

$$Q_\tau(\hat{\varepsilon}_{it}(\tau)) = \nu_i + \phi_i^{'}(\tau)x_{it} + \eta_{it}(\tau), \tag{12}$$

However, deriving an appropriate test statistic for this hypothesis is beyond the scope of this thesis.

4 Monte Carlo Simulation

4.1 Data Generating Processes

The first part of the data used in this thesis comes from a Monte Carlo simulation setup. I use this to (i) compare the 'hit rate' of the various group number selection procedures and (ii) assess the impact under- or overestimating the total number of groups has on coefficient estimates in a finite-sample setting. The first four DGPs I use to simulate the data are the same as those used in Leng et al. (2021):

DGP.1:

$$y_{it} = \alpha_{h_i} + \lambda_t + \beta_{g_i}x_{it} + (1 + \psi x_{it})\varepsilon_{it}, \quad h_i = 1, ..., 4; \; g_i = 1, 2. \tag{13}$$

Here, $\psi = 0.5$, $\lambda_t \sim Uniform(0,1)$, $x_{it} = 0.3(\alpha_{h_i} + \lambda_t) + z_{it}$, where z_{it} is i.i.d as χ_5^2, and ε_{it} is i.i.d as $N(0,1)$.

Note that there are 4 cross-section fixed effect groups and 2 slope groups in this DGP. In this setup, the first $N/4$ units have $h_i = 1$, the next $N/4$ units have $h_i = 2$, and so forth, and the units in the first two cross-section fixed effect groups have $g_i = 1$, whereas the units in the last two cross-section fixed effect groups have $g_i = 2$.

Finally, the group-specific coefficients in this DGP are as follows:

$$(\alpha_1, \alpha_2, \alpha_3, \alpha_4) = (-5, -2.5, 2.5, 5); \quad (\beta_1, \beta_2) = (-0.75, 0, 75).$$

DGP.2: This DGP is the same as DGP.1, with the exception of the ε_{it} term, which is now distributed as:

$$\varepsilon_{it} \sim \begin{cases} \text{i.i.d. } N(0,1) & \text{if } g_i = 1 \\ \text{i.i.d. Weibull(3,1)} - E[\text{Weibull}(3,1)] & \text{if } g_i = 2. \end{cases} \tag{14}$$

In other words, the units in the first slope coefficient group have ε_{it} distributed as a centered Weibull distribution with shape parameter 3 and scale parameter 1. This setting ensures that the two slope groups have differently shaped slope distributions, rather than just different slope means.

DGP.3: This DGP is the same as DGP.1, with the exception of the group assignment. In this DGP, H-group assignment remains the same as in DGP.1, but the first 3/8 of the units are assigned $g_i = 1$, while the remaining units are assigned $g_i = 2$, resulting in non-nested grouping.

DGP.4: This combines DGP.2 and DGP.3 such that group assignment follows DGP.3, the distribution of ε_{it} follows DGP.2, and the rest of the specification follows DGP.1.

The final two DGPs are new additions to this thesis.

DGP.5: This DGP is meant to induce lower levels of group separation at the tail quantiles, by increasing ψ. This results in the following model:

$$y_{it} = \alpha_{h_i} + \lambda_t + \beta_{g_i} x_{it} + (1 + x_{it})\varepsilon_{it}, \quad h_i = 1, ..., 2; \; g_i = 1, ..., 4. \tag{15}$$

In other words, ψ is increased to 1. As a result of the increased influence of the error term on the slope, there should be relatively strong group separation at the median, and weaker group separation at the tail quantiles due to random contamination.

In this model, the first $N/2$ units have $h_i = 1$ and the remainder have $h_i = 2$, while the first

$N/4$ units have $g_i = 1$, the next $N/4$ units have $g_i = 2$, and so forth, and the coefficients satisfy

$$(\alpha_1, \alpha_2) = (-5, 5); \quad (\beta_1, \beta_2, \beta_3, \beta_4) = (-1.25, -0.5, 0.5, 1.25).$$

DGP.6: The final DGP is a variation on DGP.1, where instead of four fixed effect groups and two slope groups, there are two fixed effect groups and four slope groups. This is to more fully assess the performance of the group number selection methods in selecting the right number of slope groups. The functional form of the model is the same as in DGP.1, i.e.

$$y_{it} = \alpha_{h_i} + \lambda_t + \beta_{g_i} x_{it} + (1 + \psi x_{it})\varepsilon_{it}, \quad h_i = 1, ..., 2;\ g_i = 1, ..., 4. \tag{16}$$

Here, $\psi = 0.5$, as in DGP.1. Now the first $N/2$ units have $h_i = 1$ and the remainder have $h_i = 2$, whereas the first $N/4$ units have $g_i = 1$, the next $N/4$ units have $g_i = 2$, and so forth. The model coefficients are

$$(\alpha_1, \alpha_2) = (-3.75, 3.75); \quad (\beta_1, \beta_2, \beta_3, \beta_4) = (-2.25, -0.75, 0.75, 2.25).$$

For the simulation studies, I simulate 25 replications each of all six DGPs, with (N,T) = (80, 20) and (N,T) = (160, 40) respectively, for 25 replications of 12 simulation settings in total.

4.2 Evaluation Criteria

I validate the convergence of Algorithm 1 to the global minimum - and thus the validity of our estimates - by first assessing the sample slope bias and Root Mean Square Error (RMSE) for $\tau = 0.1$, 0.3, 0.5, 0.7, and 0.9, as well as the Misclustering Frequency (MF). I follow Leng et al. (2021) in calculating three measures of the MF: the overall MF, the G-specific MF, and the H-specific MF.

The slope bias is calculated as

$$Bias(\tau) = (1/GR) \sum_{r=1}^{R} \sum_{g=1}^{G} (\hat{\beta}_g(\tau) - \beta_g^{true}(\tau)),$$

whereas the RMSE is calculated as

$$RMSE(\tau) = (1/R) \sum_{r=1}^{R} \sqrt{(1/G) \sum_{g=1}^{G} (\hat{\beta}_g(\tau) - \beta_g^{true}(\tau))^2}.$$

In these equations, R denotes the number of replications; thus, $R = 25$.

Moreover, I calculate the overall clustering misfrequency as

$$MF_{overall} = (1/R) \sum_{r=1}^{R} \big(1 - (1/N) \sum_{i=1}^{N} I(\hat{g}_i = g_i^{true}, \hat{h}_i = h_i^{true})\big),$$

the G-specific clustering misfrequency as

$$MF_G = (1/R) \sum_{r=1}^{R} \big(1 - (1/N) \sum_{i=1}^{N} I(\hat{g}_i = g_i^{true})\big),$$

and the H-specific clustering misfrequency as

$$MF_H = (1/R) \sum_{r=1}^{R} \big(1 - (1/N) \sum_{i=1}^{N} I(\hat{h}_i = h_i^{true})\big).$$

For the primary purpose of this simulation study, I additionally calculate the sample probabilities to select certain levels of G and H for each group number selection procedure discussed in the preceding sections, across the 25 replications. These measures are simply calculated as:

$$p_m(G) = (1/R) \sum_{r=1}^{R} I\big(G_{rm} = G\big) \text{ and } p_m(H) = (1/R) \sum_{r=1}^{R} I\big(H_{rm} = H\big),$$

where $p_m(G)$ and $p_m(H)$ denote the sample probabilities of model m selecting G slope groups and H fixed effect groups, respectively, and G_{rm} and H_{rm} denote the numbers of slope and fixed effect groups, respectively, chosen by procedure m in replication r.

I conclude by assessing the most common *incorrectly* estimated levels of G and H, and gauging the average individual-level bias and RMSE as a result of this over- or under-estimation. I calculate this individual-level bias as

$$Bias_{(G,H)\neq(G^{true},H^{true})}(\tau) = (1/NR) \sum_{r=1}^{R} \sum_{i=1}^{N} \big(\hat{\beta}_{\hat{g}_i}(\tau) - \beta_{g_i}(\tau)\big),$$

Table 1: Misclustering Frequencies given G = 2 and H = 4

	N = 80, T = 20	N = 160, T = 40	N = 80, T = 20	N = 160, T = 40
	DGP.1		DGP.2	
$MF_{overall}$	0.041	0.034	0.010	0.013
MF_G	0.001	0.000	0.000	0.000
MF_H	0.041	0.034	0.010	0.013
	DGP.3		DGP.4	
$MF_{overall}$	0.052	0.010	0.004	0.001
MF_G	0.002	0.000	0.002	0.001
MF_H	0.052	0.010	0.004	0.001
	DGP.5		DGP.6	
$MF_{overall}$	0.186	0.062	0.003	0.000
MF_G	0.186	0.062	0.003	0.000
MF_H	0.001	0.000	0.003	0.000

and this individual-level RMSE as

$$RMSE_{(G,H)\neq(G^{true},H^{true})}(\tau) = (1/R)\sum_{r=1}^{R}\sqrt{(1/N)\sum_{i=1}^{N}\left(\hat{\beta}_{\hat{g}_i}(\tau) - \beta_{g_i}(\tau)\right)^2}.$$

4.3 Simulation Results

4.3.1 Model Performance Given G and H

Table 1 displays the misclustering frequencies for the given combinations of DGP and sample size, under the assumption that G and H are known (to be 2 and 4, respectively, for DGP.1-DGP.4, and 4 and 2, respectively, for DGP.5-DGP.6). These results suggest that Algorithm 1 converged to the global, rather than the local, solution most or all of the time. The slope classification is performed very accurately for all DGPs and sample sizes except the smaller sample of DGP.5, whereas the fixed effect classification is accurate across all settings. Although the slope group classification performance for DGP.5 seems poor, it is relatively accurate when taking into account the weak group separation induced in this DGP.

In addition, Table 2 displays the RMSE and Bias incurred in the estimation of all of these models, for various levels of tau and different sample sizes. The results indicate small sample biases, once again suggesting that Algorithm 1 converged to a global minimum. However, the results also display relatively large RMSEs for DGP.2, DGP.4, and DGP.5. This can be explained by estimation uncertainty induced by the heterogeneous slope distributions and weak group

Table 2: Bias and RMSE per level of tau and sample size, given G = 2 and H = 4

		N = 80, T = 20		N = 160, T = 40			N = 80, T = 20		N = 160, T = 40	
τ		Bias	RMSE	Bias	RMSE		Bias	RMSE	Bias	RMSE
0.1	DGP.1	0.013	0.085	0.018	0.049	DGP.2	0.007	0.205	0.006	0.209
0.3		0.003	0.052	0.016	0.044		-0.037	0.093	-0.032	0.083
0.5		0.001	0.047	0.006	0.041		0.001	0.045	0.006	0.025
0.7		-0.005	0.050	0.009	0.038		0.041	0.104	0.038	0.103
0.9		-0.016	0.079	-0.004	0.044		-0.020	0.206	-0.007	0.217
0.1	DGP.3	0.030	0.076	0.011	0.037	DGP.4	-0.099	0.278	-0.109	0.282
0.3		0.011	0.051	0.004	0.029		-0.067	0.127	-0.068	0.125
0.5		0.008	0.050	0.003	0.026		-0.002	0.030	0.001	0.016
0.7		0.001	0.050	0.003	0.028		0.063	0.125	0.059	0.120
0.9		-0.022	0.071	-0.010	0.035		0.081	0.262	0.089	0.267
0.1	DGP.5	0.054	0.222	0.034	0.088	DGP.6	0.021	0.113	-0.002	0.052
0.3		0.021	0.160	0.010	0.072		0.009	0.068	-0.005	0.036
0.5		-0.029	0.152	-0.005	0.071		0.004	0.067	-0.005	0.033
0.7		-0.064	0.160	-0.015	0.076		-0.007	0.075	0.001	0.032
0.9		-0.108	0.197	-0.038	0.091		-0.021	0.090	-0.012	0.048

separation in these DGPs.

Taken together, these results suggest that, at least given the correct G and H, Algorithm 1 yields accurate estimates for both the group membership parameters as well as the regression coefficients.

4.3.2 Group Number Selection Performance

I now turn to the main results of this thesis: the comparison of the group number selection performance of the four different procedures discussed in the preceding sections. Table 3 displays the sample probability of selecting any level of G and H for all four methods, both sample sizes and all six considered DGPs. One interesting observation is that, while for $(N, T) = (80, 20)$ the Max-IC procedure appears to generally be the best choice to select G and H, this advantage shrinks significantly for $(N, T) = (160, 40)$, as the composite IC procedure's relative performance appears to scale much better with sample size. This result suggests that the Max-IC procedure has an edge for small sample sizes, but loses this edge for moderate to large sample sizes. In this assessment, it is important to note that the Max-IC procedure underestimates the number of groups much less often than the other methods, though it tends to overestimate. This latter occurrence is illustrated most clearly in the results for DGP.3 with $(N, T) = (160, 40)$, where Max-IC does not predict the correct value of H a single time, overestimating it in each replication. The relative strength of the Max-IC procedure, on the other hand, is most clearly displayed in the results for DGP.5 with $(N,T) = (80,20)$, where it never underestimates H, and only underestimates G 28% of the time, whereas the WIC procedure underestimates G 80% of the time, and the composite IC and testing procedures even do so 96% and 100% of the time, respectively. However, for the larger sample of DGP.5, the composite IC procedure achieves perfect classification, something the Max-IC procedure fails to do.

An additional notable observation is that the WIC procedure displays very unstable behavior, often producing far-off estimates. The WIC is the only IC method that under- or overestimated a group number by more than one, doing so for every single DGP. In addition, while the WIC procedure produced relatively good estimates for DGP.3 with the smaller sample size, it produced by far the worst results for the same DGP with the larger sample size. This suggests that the WIC is highly sensitive to changes in the data. For DGP.5, the DGP notable for having greater variation in group separation across the quantiles, the WIC does marginally outperform the composite IC in the small sample case but is found to perform worse than the Max-IC procedure in the same setting, and much worse than the composite IC procedure in the large sample case.

Moving on to the testing procedure, the results indicate that this procedure is only effective at identifying H for a very large sample size, and is only effective at identifying G if there is significantly large group separation or, again, sample size. A possible explanation for the former phenomenon is that moderately underestimating H does not lead to large biases in the slope coefficients, and, since the testing procedure relies on biases in the slope coefficient, the procedure fails to properly identify fixed effect group underestimation. The latter occurrence simply indicates that the testing procedure, for small and moderate sample sizes, is less capable of identifying weakly separated groups than the IC procedures. This may be the consequence of the lack of statistical power of the LM test as a result of only using one, rather than K, series of residuals.

Given these results, the approach that appears most likely to yield accurate group number estimates involves selecting the number of fixed effect groups based on the composite IC or Max-IC procedures and selecting the number of slope groups based on the LM testing or composite IC procedure if N and T are very large or there is strong group separation across the quantiles, and otherwise using the composite IC or Max-IC procedure for the estimation of G as well.

Table 3: Sample probability of selecting different Group Numbers for each method

Method	(N, T)		1	2	3	4	5	6	1	2	3	4	5	6
			DGP.1						DGP.2					
IC	(80, 20)	H	0.00	0.00	0.24	**0.76**	0.00	0.00	0.00	0.00	0.00	**1.00**	0.00	0.00
		G	0.00	**0.76**	0.24	0.00	0.00	0.00	0.00	**0.88**	0.12	0.00	0.00	0.00
Max-IC	(80, 20)	H	0.00	0.00	0.00	**1.00**	0.00	0.00	0.00	0.00	0.00	**0.96**	0.04	0.00
		G	0.00	**0.28**	0.72	0.00	0.00	0.00	0.00	**0.76**	0.24	0.00	0.00	0.00
WIC	(80, 20)	H	0.00	0.16	0.16	**0.68**	0.00	0.00	0.00	0.00	0.00	**0.80**	0.16	0.04
		G	0.00	**0.68**	0.20	0.12	0.00	0.00	0.00	**0.76**	0.24	0.00	0.00	0.00
Testing	(80, 20)	H	0.00	1.00	0.00	**0.00**	0.00	0.00	0.00	1.00	0.00	**0.00**	0.00	0.00
		G	0.00	**0.96**	0.04	0.00	0.00	0.00	0.00	**0.96**	0.04	0.00	0.00	0.00
IC	(160, 40)	H	0.00	0.00	0.16	**0.68**	0.16	0.00	0.00	0.00	0.08	**0.92**	0.00	0.00
		G	0.00	**0.76**	0.24	0.00	0.00	0.00	0.00	**0.76**	0.24	0.00	0.00	0.00
Max-IC	(160, 40)	H	0.00	0.00	0.08	**0.72**	0.20	0.00	0.00	0.00	0.08	**0.92**	0.00	0.00
		G	0.00	**0.60**	0.40	0.00	0.00	0.00	0.00	**0.72**	0.28	0.00	0.00	0.00
WIC	(160, 40)	H	0.00	0.00	0.12	**0.68**	0.20	0.00	0.00	0.00	0.20	**0.80**	0.00	0.00
		G	0.00	**0.76**	0.24	0.00	0.00	0.00	0.00	**0.68**	0.28	0.04	0.00	0.00
Testing	(160, 40)	H	0.00	0.92	0.04	**0.04**	0.00	0.00	0.00	0.92	0.00	**0.08**	0.00	0.00
		G	0.00	**0.96**	0.04	0.00	0.00	0.00	0.00	**0.96**	0.04	0.00	0.00	0.00
			DGP.3						DGP.4					
IC	(80, 20)	H	0.00	0.00	0.44	**0.56**	0.00	0.00	0.00	0.00	0.00	**1.00**	0.00	0.00
		G	0.00	**0.56**	0.44	0.00	0.00	0.00	0.00	**1.00**	0.00	0.00	0.00	0.00
Max-IC	(80, 20)	H	0.00	0.00	0.04	**0.96**	0.00	0.00	0.00	0.00	0.00	**1.00**	0.00	0.00
		G	0.00	**0.08**	0.92	0.00	0.00	0.00	0.00	**0.72**	0.28	0.00	0.00	0.00
WIC	(80, 20)	H	0.00	0.00	0.28	**0.72**	0.00	0.00	0.00	0.00	0.12	**0.64**	0.20	0.04
		G	0.00	**0.60**	0.32	0.08	0.00	0.00	0.00	**0.60**	0.36	0.04	0.00	0.00
Testing	(80, 20)	H	0.00	0.96	0.04	**0.00**	0.00	0.00	0.00	0.80	0.12	**0.08**	0.00	0.00
		G	0.00	**0.92**	0.04	0.04	0.00	0.00	0.00	**1.00**	0.00	0.00	0.00	0.00
IC	(160, 40)	H	0.00	0.00	0.04	**0.96**	0.00	0.00	0.00	0.00	0.00	**1.00**	0.00	0.00
		G	0.00	**0.40**	0.60	0.00	0.00	0.00	0.00	**0.84**	0.12	0.04	0.00	0.00
Max-IC	(160, 40)	H	0.00	0.00	0.00	**0.96**	0.04	0.00	0.00	0.00	0.00	**0.72**	0.28	0.00
		G	0.00	**0.00**	0.96	0.04	0.00	0.00	0.00	**0.28**	0.64	0.08	0.00	0.00
WIC	(160, 40)	H	0.00	0.00	0.36	**0.44**	0.20	0.00	0.00	0.00	0.04	**0.68**	0.28	0.00
		G	0.00	**0.08**	0.60	0.32	0.00	0.00	0.00	**0.36**	0.44	0.20	0.00	0.00
Testing	(160, 40)	H	0.00	0.92	0.04	**0.04**	0.00	0.00	0.00	0.36	0.28	**0.36**	0.00	0.00
		G	0.00	**0.92**	0.08	0.00	0.00	0.00	0.00	**0.96**	0.04	0.00	0.00	0.00
			DGP.5						DGP.6					
IC	(80, 20)	H	0.00	**1.00**	0.00	0.00	0.00	0.00	0.00	**1.00**	0.00	0.00	0.00	0.00
		G	0.00	0.00	0.96	**0.04**	0.00	0.00	0.00	0.00	0.00	**1.00**	0.00	0.00
Max-IC	(80, 20)	H	0.00	**0.56**	0.44	0.00	0.00	0.00	0.00	**1.00**	0.00	0.00	0.00	0.00
		G	0.00	0.00	0.28	**0.72**	0.00	0.00	0.00	0.00	0.00	**1.00**	0.00	0.00
WIC	(80, 20)	H	0.00	**0.80**	0.20	0.00	0.00	0.00	0.00	**0.84**	0.16	0.00	0.00	0.00
		G	0.00	0.04	0.76	**0.20**	0.00	0.00	0.00	0.00	0.00	**0.72**	0.28	0.00
Testing	(80, 20)	H	0.00	**1.00**	0.00	0.00	0.00	0.00	0.00	**1.00**	0.00	0.00	0.00	0.00
		G	0.00	0.52	0.48	**0.00**	0.00	0.00	0.00	0.00	0.00	**1.00**	0.00	0.00
IC	(160, 40)	H	0.00	**1.00**	0.00	0.00	0.00	0.00	0.00	**0.96**	0.04	0.00	0.00	0.00
		G	0.00	0.00	0.00	**1.00**	0.00	0.00	0.00	0.00	0.00	**1.00**	0.00	0.00
Max-IC	(160, 40)	H	0.00	**0.76**	0.24	0.00	0.00	0.00	0.00	**0.64**	0.32	0.00	0.00	0.00
		G	0.00	0.00	0.00	**0.88**	0.12	0.00	0.00	0.00	0.00	**0.84**	0.16	0.00
WIC	(160, 40)	H	0.00	**0.72**	0.28	0.00	0.00	0.00	0.00	**0.84**	0.16	0.00	0.00	0.00
		G	0.00	0.00	0.12	**0.76**	0.12	0.00	0.00	0.00	0.04	**0.56**	0.32	0.08
Testing	(160, 40)	H	0.00	**1.00**	0.00	0.00	0.00	0.00	0.00	**1.00**	0.00	0.00	0.00	0.00
		G	0.00	0.20	0.56	**0.24**	0.00	0.00	0.00	0.00	0.00	**1.00**	0.00	0.00

Note: Bold numbers indicate correct group number estimates.

4.3.3 Coefficient Impact of Incorrect Group Number Estimation

Table 4: Individual Bias and RMSE per level of tau and sample size, for various group number misclassifications

			N = 80, T = 20		N = 160, T = 40				N = 80, T = 20		N = 160, T = 40	
τ	(G, H)	DGP	Bias	RMSE	Bias	RMSE	(G, H)	DGP	Bias	RMSE	Bias	RMSE
0.1	(3, 3)	1	0.049	0.182	0.050	0.152	(3, 4)	1	0.051	0.163	0.028	0.116
0.3			0.022	0.175	0.032	0.149			0.021	0.160	0.019	0.112
0.5			0.013	0.166	0.016	0.145			0.009	0.163	0.012	0.109
0.7			-0.011	0.168	0.007	0.142			0.000	0.170	0.004	0.106
0.9			-0.046	0.181	-0.014	0.141			-0.050	0.183	-0.016	0.106
0.1	(2, 3)	1	0.090	0.150	0.052	0.074	(3,2)	6	0.088	0.592	0.106	0.564
0.3			0.063	0.116	0.046	0.067			0.094	0.599	0.129	0.586
0.5			0.050	0.102	0.047	0.068			0.108	0.614	0.143	0.607
0.7			0.036	0.098	0.035	0.062			0.104	0.628	0.153	0.636
0.9			-0.011	0.111	0.023	0.059			0.066	0.650	0.120	0.674

In this section, I assess the impact that incorrect estimation of the group numbers has on slope coefficient estimates. To this end, one of the most common incorrect estimations of the group numbers among the replications in the preceding simulation study is $(G, H) = (3, 3)$ for the first four DGPs. This involves an incorrect estimation in both dimensions: the underestimation of G and the overestimation of H. Two other common incorrect estimates for the first four DGPs are $(G, H) = (3, 4)$, involving only the overestimation of G, and $(G, H) = (2, 3)$, involving only the underestimation of H. Finally, G is often underestimated at a value of 3 in DGP.5 and DGP.6, and so I also assess $(G, H) = (3, 2)$. To keep matters simple, I only assess the impact of the former three incorrect estimations for DGP.1 - the nested model with basic error terms and $(G_{true}, H_{true}) = (2, 4)$ -, and the impact of the last incorrect estimation for DGP.6 - the nested model with basic error terms and $(G_{true}, H_{true}) = (4, 2)$. Overestimation of H should have only a very small impact on estimated slope coefficients, and the preceding analysis suggests that it does not occur often. Therefore, I do not assess the impact of H overestimation. The results of the analyses of this section can be found in Table 4.

One notable observation is that the sample biases and RMSEs appear to be much higher than the biases and RMSEs in Table 2 across the board in relative terms, but, with the exception of the case where G is underestimated, they are not very large in absolute terms. This is an interesting observation because the individual biases and RMSEs displayed in Table 4 also incorporate error due to misclustering, whereas the figures in Table 2 only take into account error in the coefficient estimation. This, therefore, suggests that overestimating G or underestimating H by 1 only has a moderately adverse effect on individuals' slope estimates.

A second notable observation is that $(G, H) = (3, 3)$ and $(G, H) = (3, 4)$ yield very similar biases and RMSEs, while the latter features both overestimation of G as well as underestimation of H, whereas the former features only overestimation of G. Given that $(G, H) = (2, 3)$ results in notably higher bias figures, this suggests that the overestimation of G can 'compensate' for the underestimation of H when it comes to individual-level slope bias; at least in some datasets. However, underestimating one number of groups without overestimating the other, as with $(G, H) = (2, 3)$ for DGP.1, yields relatively large biases, especially for small sample sizes.

Finally, the biases and, in particular, the RMSEs recorded with the underestimation of G are very large. Moreover, they appear to become worse rather than better with the larger sample size. Comparing the results of overestimating G by one in DGP.1, to the results of underestimating G by one in DGP.6, the case of underestimation yields biases many orders of magnitude larger, and RMSEs that are roughly four times as large as well.

In sum, these results suggest that, also in finite-sample studies, overestimation of the group numbers is highly preferable to underestimation, strengthening the argument in favor of the Max-IC procedure, which only rarely underestimates the group numbers.

5 State Productivity Data Application

Finally, I apply all four group number selection procedures discussed in the prior sections to real data on productivity, the level of public and private capital, non-agricultural labor input, as well as unemployment rates of 48 US states over the period 1970-1986. As the data is collected annually, the dataset has $N = 48$ units and $T = 17$ time periods. This dataset was first considered in Munnell et al. (1990) and was also extensively discussed in Baltagi (2008) and a wide array of other papers over the years. The key question in these studies is how elastic productivity is with respect to public capital; a question highly relevant to the US in recent decades, as critics point to decaying infrastructure in the US as a detractor of economic growth (see e.g. Aschauer (1989)). However, this problem has typically been approached using mean-based methods, while there may in practice be differences across the quantiles. Significant differences across quantiles in production functions have been found before, e.g. by Bernini et al. (2004) and Nyamekye et al. (2016). Moreover, two-dimensional heterogeneity, as in the models utilized in this thesis, has not been applied to this problem before. Typically, only heterogeneity in the intercept is considered in these studies. Therefore, I propose to estimate the states' productivity using the following Cobb-Douglas model:

$$Q_\tau(Y_{it}|x_{it}) = \alpha_{h_i}(\tau) + \lambda_t(\tau) + \beta_{1,g_i}(\tau)pubc_{it} + \beta_{2,g_i}(\tau)privc_{it} + \beta_{3,g_i}(\tau)labor_{it} + \beta_{4,g_i}(\tau)unemp_{it}, \tag{17}$$

where Y_{it} denotes the log gross product of state i at time t, $pubc_{it}$ denotes the natural logarithm of the public capital of state i at time t, $privc_{it}$ denotes the natural logarithm of the private capital of state i at time t, $labor_{it}$ denotes the natural logarithm of the labor input (minus agricultural labor) of state i at time t, and $unemp_{it}$ denotes the unemployment of state i at time t as a percentage. Our main parameters of interest are the $\beta_{1,g_i}(\tau)$, which represent the group-specific elasticity of aggregate state production with respect to public capital, at the τ'th quantile. As before, I consider $\tau_k = \frac{k}{K+1}$ for $k \in 1, ..., K = 9$.

Applying all four procedures of determining G and H considered in this thesis to the state productivity data reveals that the IC procedures all agree on $(G, H) = (1, 6)$, with $(G, H) = (2, 5)$ yielding only slightly higher composite and weighted IC values. A probable reason for the low number of estimated G groups is that some of the four explanatory variables do not display grouped patterns, even if others may, or the slope coefficients are simply homogeneous across all units.

This latter hypothesis is not supported, however, by the testing procedure, which rejects $(G, H) = (1, 1)$ but fails to reject both $(G, H) = (1, 2)$ as well as $(G, H) = (2, 1)$. Given that both of these combinations have a total group number of three, I let the magnitude of the test statistic be the deciding factor (with smaller in size being preferred), resulting in an estimated group number combination of $(G, H) = (2, 1)$. Incidentally, this combination also boasts a lower objective value. However, as also with $(G, H) = (1, 2)$ the median regression residuals appear to be uncorrelated with the regressors, and bearing in mind that the testing procedure tends to underestimate H, this result also lends some credence to the information criterion estimates of $(G, H) = (1, 6)$. In any case, as the methods do not agree, I will estimate the model for three different group number combinations: the IC estimates $(1, 6)$, the testing estimate $(2, 1)$, and a combination of the two, $(2, 6)$. This is in line with my previous assessment that choosing the number of slope groups based on the testing procedure and the number of fixed effect groups based on the IC procedure can yield better results than the separate estimates, if N and T are large (which is not satisfied in this case) or if there is sufficient G-group separation. The results of the quantile regression for all three combinations of G and H can be found in Table 5, below.

There are several noteworthy results in this table. For one, when considering the results for $(G, H) = (2, 1)$ and $(G, H) = (2, 6)$, the difference in coefficient estimates between the two groups tends to be much larger than their respective standard errors, suggesting that there are indeed significant slope differences between the groups. Moreover, there does not appear to be much difference between coefficient estimates across the different quantiles. In the regression with homogeneous slope coefficients, i.e., $(G, H) = (1, 6)$, only the effect of the level of public capital on productivity appears to decrease in an economically significant manner as the quantile increases, but this effect remains relatively stable across the quantiles when including an additional slope coefficient group.

In addition, the estimated coefficients of the two slope coefficient groups appear to change significantly with the inclusion of heterogeneous fixed effects. The difference in the estimated elasticity of productivity with respect to public capital between groups 1 and 2 is rather small with homogeneous fixed effects, i.e. if $(G, H) = (2, 1)$, but quite large with heterogeneous fixed effects, i.e. if $(G, H) = (2, 6)$. At the same time, the standard errors of the slope coefficients decrease significantly with the inclusion of heterogeneous intercepts. These results suggest that the model with heterogeneous fixed effects is the more appropriate of the two.

On the other hand, one can also consider the difference between the estimated coefficients for the model with homogeneous slope coefficients but heterogeneous fixed effects, $(G, H) = (1, 6)$, and the model with heterogeneous slope and fixed effect coefficients, $(G, H) = (2, 6)$. The estimated homogeneous effect of public capital is much lower than the estimated effect of public capital of group 2 in the heterogeneous slope model. Similarly, the estimated homogeneous effect of private capital is much larger than the estimated effect of private capital of group 2 in the larger model. The estimated homogeneous effect of labor input, on the other hand, is relatively similar to the estimated effect of both groups in the larger model, suggesting that there is no heterogeneity in this coefficient.

In sum, these results suggest that there is significant elasticity of productivity of states with respect to public capital, in the direction that one would expect: an increase in public capital is associated with increased productivity. However, these effects do not appear homogeneous

Table 5: State productivity quantile regression estimates for various group numbers

		$\tau =$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
		(G, H) = (1,6)									
pubc	Group 1		0.141	0.119	0.122	0.117	0.104	0.094	0.067	0.057	0.077
			(0.026)	(0.011)	(0.010)	(0.013)	(0.013)	(0.012)	(0.019)	(0.024)	(0.020)
	Group 2										
privc	Group 1		0.360	0.356	0.360	0.360	0.366	0.373	0.380	0.375	0.380
			(0.016)	(0.007)	(0.006)	(0.008)	(0.008)	(0.008)	(0.012)	(0.015)	(0.013)
	Group 2										
labor	Group 1		0.578	0.605	0.597	0.602	0.605	0.606	0.624	0.633	0.604
			(0.012)	(0.009)	(0.008)	(0.010)	(0.010)	(0.009)	(0.015)	(0.019)	(0.016)
	Group 2										
unemp	Group 1		-0.014	-0.011	-0.011	-0.011	-0.010	-0.009	-0.009	-0.010	-0.012
			(0.003)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
	Group 2										
		(G, H) = (2,1)									
pubc	Group 1		0.277	0.219	0.189	0.199	0.216	0.215	0.230	0.224	0.219
			(0.023)	(0.014)	(0.013)	(0.015)	(0.014)	(0.015)	(0.018)	(0.015)	(0.011)
	Group 2		0.244	0.209	0.220	0.251	0.271	0.253	0.243	0.200	0.176
			(0.034)	(0.020)	(0.019)	(0.022)	(0.021)	(0.022)	(0.026)	(0.022)	(0.016)
privc	Group 1		0.145	0.182	0.209	0.210	0.213	0.211	0.208	0.202	0.203
			(0.019)	(0.011)	(0.010)	(0.012)	(0.011)	(0.012)	(0.014)	(0.012)	(0.009)
	Group 2		0.260	0.276	0.270	0.264	0.269	0.277	0.284	0.294	0.309
			(0.017)	(0.010)	(0.009)	(0.011)	(0.010)	(0.011)	(0.013)	(0.011)	(0.008)
labor	Group 1		0.617	0.642	0.643	0.634	0.612	0.612	0.598	0.606	0.607
			(0.017)	(0.010)	(0.009)	(0.011)	(0.010)	(0.011)	(0.013)	(0.011)	(0.008)
	Group 2		0.525	0.544	0.537	0.508	0.481	0.490	0.495	0.534	0.539
			(0.025)	(0.015)	(0.013)	(0.016)	(0.015)	(0.016)	(0.019)	(0.016)	(0.011)
unemp	Group 1		0.004	0.000	-0.002	-0.004	-0.004	-0.002	-0.002	-0.007	-0.006
			(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)
	Group 2		-0.010	-0.010	-0.012	-0.014	-0.015	-0.014	-0.013	-0.020	-0.016
			(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)
		(G, H) = (2,6)									
pubc	Group 1		0.075	0.094	0.103	0.095	0.092	0.079	0.084	0.084	0.084
			(0.009)	(0.007)	(0.009)	(0.009)	(0.009)	(0.011)	(0.007)	(0.010)	(0.015)
	Group 2		0.341	0.337	0.337	0.333	0.333	0.330	0.336	0.342	0.344
			(0.011)	(0.008)	(0.010)	(0.010)	(0.010)	(0.013)	(0.008)	(0.011)	(0.017)
privc	Group 1		0.314	0.304	0.300	0.303	0.310	0.319	0.320	0.320	0.318
			(0.007)	(0.005)	(0.007)	(0.006)	(0.006)	(0.008)	(0.005)	(0.007)	(0.011)
	Group 2		0.197	0.197	0.202	0.211	0.214	0.228	0.229	0.231	0.235
			(0.007)	(0.005)	(0.006)	(0.006)	(0.006)	(0.008)	(0.005)	(0.007)	(0.011)
labor	Group 1		0.679	0.665	0.662	0.666	0.663	0.666	0.661	0.659	0.662
			(0.007)	(0.005)	(0.007)	(0.006)	(0.007)	(0.008)	(0.005)	(0.007)	(0.011)
	Group 2		0.500	0.501	0.495	0.487	0.482	0.467	0.460	0.451	0.441
			(0.008)	(0.006)	(0.008)	(0.007)	(0.007)	(0.009)	(0.006)	(0.008)	(0.013)
unemp	Group 1		-0.003	-0.004	-0.004	-0.004	-0.005	-0.004	-0.006	-0.006	-0.006
			(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
	Group 2		-0.011	-0.011	-0.011	-0.010	-0.010	-0.010	-0.012	-0.013	-0.013
			(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)

Note: Standard errors, based on the assumption of i.i.d. DGP errors, are given in parentheses.

across all states. In the context of the model with $(G, H) = (2, 6)$, the states of the US can be partitioned into two groups. The states of group 1 display a relatively low elasticity and include Alabama, Arizona, Arkansas, California, Connecticut, Georgia, Idaho, Iowa, Massachusetts, Minnesota, Mississippi, Missouri, Montana, Nebraska, Nevada, New Hampshire, New Jersey, New Mexico, North Dakota, South Dakota, Tennessee, Texas, Utah, Vermont, Virginia, Washington, West Virginia, and Wisconsin. The states of group 2, on the other hand, display a large elasticity of production with respect to public capital and include Colorado, Delaware, Florida, Illinois, Indiana, Kansas, Kentucky, Louisiana, Maine, Maryland, Michigan, New York, North Carolina, Ohio, Oklahoma, Oregon, Pennsylvania, Rhode Island, South Carolina, and Wyoming. As this classification relies on many hidden characteristics of the states, it is not clear what separates these two groups of states; it seems, however, that it is not a political, ideological, or geographic divide.

At this point, it is useful to consider the extensive existing literature on estimating the elasticity of productivity with respect to public capital in the US. Most of these studies have used the same or a similar dataset as I, and their results are therefore easily compared to the results in this section. The literature produced two contradicting results: Aschauer (1989), Munnell et al. (1990), and Cook et al. (1990) find large, significantly positive elasticities of productivity with respect to public capital ranging from 0.06 to 0.39, whereas Evans and Karras (1994), Eisner et al. (1991), and Holtz-Eakin et al. (1994) find insignificant or even significantly negative elasticities. The primary difference between these studies is that the latter batch controlled for state-specific effects, whereas the former neglected to do so. However, the results in this thesis suggest that when assuming homogeneity across the slope coefficients, six homogeneous fixed effect groups are sufficient to model the data, and the associated estimated elasticities range from 0.057 near the top quantile to 0.141 at the bottom quantile. These estimates are therefore at the lower end of the range found in the former set of studies, but significantly higher than the results found in the latter set of studies. In addition, it is striking that, when accounting for both fixed effect as well as slope heterogeneity, the estimated elasticities of group 2 ranging from 0.330 to 0.341 strongly resemble most estimated elasticities found in the former set of papers, whereas the estimated elasticities for group 1 ranging from 0.075 to 0.103 are more in line with the latter set of papers.

To also account for possible underestimation of the number of fixed effect groups in my models, I moreover estimate the regression with state-specific fixed effects and two slope coefficient groups. The results of this regression may be found in Table 6. This regression yields estimated elasticities of productivity with respect to public capital between -0.13 and -0.04 for the states of group 1, the composition of which remains largely the same as before, and estimated elasticities of productivity with respect to public capital between -0.04 and 0.18 for the states of group 2. The estimated elasticities for the states of group 1, therefore, strongly resemble the results of Evans and Karras (1994), Eisner et al. (1991), and Holtz-Eakin et al. (1994), whereas the estimated elasticities for the states of group 2 strongly resemble the results of Munnell et al. (1990). Taken together, the results of this section suggest three key takeaways: (i) that there is group-wise heterogeneity at play in the elasticity of US state productivity with respect to public capital and a failure to account for this heterogeneity may result in biased results, (ii) that the elasticity of production with respect to public capital is somewhat heterogeneous across the quantiles of state production, but not in an economically significant manner, and finally, (iii) that the choice of group number selection procedure can have significant consequences for coefficient estimates in practical studies.

Table 6: State productivity quantile regression estimates with state-fixed effects and 2 slope groups

		$\tau =$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
		(G, H) = (2, 48)									
pubc	Group 1		-0.055	-0.098	-0.109	-0.096	-0.106	-0.125	-0.077	-0.074	-0.042
			(0.021)	(0.026)	(0.026)	(0.029)	(0.029)	(0.029)	(0.025)	(0.029)	(0.024)
	Group 2		-0.044	0.030	0.044	0.043	0.092	0.096	0.074	0.130	0.176
			(0.027)	(0.035)	(0.033)	(0.039)	(0.039)	(0.037)	(0.033)	(0.037)	(0.032)
privc	Group 1		0.104	0.089	0.066	0.075	0.057	0.091	0.136	0.123	0.123
			(0.021)	(0.027)	(0.026)	(0.030)	(0.030)	(0.029)	(0.025)	(0.029)	(0.025)
	Group 2		0.054	0.015	0.037	0.140	0.211	0.235	0.234	0.277	0.302
			(0.024)	(0.031)	(0.030)	(0.034)	(0.034)	(0.030)	(0.029)	(0.033)	(0.029)
labor	Group 1		0.824	0.863	0.875	0.884	0.898	0.899	0.879	0.887	0.904
			(0.021)	(0.027)	(0.026)	(0.030)	(0.030)	(0.029)	(0.025)	(0.029)	(0.025)
	Group 2		1.003	0.969	0.902	0.823	0.701	0.679	0.751	0.646	0.609
			(0.029)	(0.037)	(0.035)	(0.041)	(0.041)	(0.040)	(0.035)	(0.039)	(0.034)
unemp	Group 1		-0.003	-0.001	-0.001	0.001	0.000	0.000	-0.001	-0.001	-0.001
			(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
	Group 2		-0.002	-0.002	-0.002	-0.003	-0.006	-0.005	-0.004	-0.006	-0.006
			(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)

Note: Standard errors, based on the assumption of i.i.d. DGP errors, are given in parentheses.

6 Conclusion and Discussion

In this thesis, I introduced two new procedures, the Weighted IC and testing procedures, to select the number of slope and fixed effect groups in a two-dimensionally grouped quantile regression setting. Moreover, I compared the performance of these methods to the pre-existing composite IC and Max-IC procedures in a Monte Carlo simulation study. The results of this comparison suggest that the composite IC and max-IC generally outperform both alternative procedures and that the max-IC procedure is often a safer choice than the composite IC procedure, especially for smaller samples.

In addition, I applied all four methods to a dataset on the productivity, public and private capital, unemployment, and labor input of 48 states over a 17-year period, which I used to investigate the elasticity of productivity with respect to public capital. I found that all IC methods agree on homogeneous slope coefficients, while the testing procedure suggests heterogeneous slope coefficients, indicating that the choice of group number selection procedure can have significant practical consequences. This analysis also yielded evidence for heterogeneity in the elasticity of state productivity with respect to public capital, which has been neglected in earlier studies.

There are, of course, several parts of this thesis that warrant future attention. For one, the weighted IC procedure proposed in 3.2.2 yielded unstable results, and alternative weighted IC approaches may yield more stable and accurate results. For example, the logarithmic transformation in the IC may be removed to accommodate the direct weighting of residuals, accompanied with an appropriate adjustment of the tuning parameter. I suggest exploring such adjustments in future work to more fully assess the feasibility of a weighted IC procedure.

In addition, the testing procedure outlined in section 3.3 can be expanded in line with my derivations up to (12), which I expect to yield accurate estimations for much smaller values of N and T due to the much larger amount of data available, as well as more robust results as they do not depend solely on the behavior of the DGP at the median.

As a final word of consideration, it would perhaps be productive for future simulation studies to utilize more replications than I have maintained in this study, which was limited partially by time constraints. Replications in the order of 100 or 1000 would allow for more definitive conclusions to be drawn from the results.

References

Angrist, J., Chernozhukov, V., and Fernández-Val, I. (2006). Quantile regression under misspecification, with an application to the us wage structure. *Econometrica*, 74(2):539–563.

Aschauer, D. A. (1989). Is public expenditure productive? *Journal of monetary economics*, 23(2):177–200.

Baltagi, B. H. (2008). *Econometric analysis of panel data*, volume 4. Springer.

Bernini, C., Freo, M., and Gardini, A. (2004). Quantile estimation of frontier production function. *Empirical Economics*, 29(2):373–381.

Bonhomme, S. and Manresa, E. (2015). Grouped patterns of heterogeneity in panel data. *Econometrica*, 83(3):1147–1184.

Cook, L. M., Munnell, A. H., et al. (1990). How does public infrastructure affect regional economic performance? *New England economic review*, (Sep):11–33.

Eisner, R. et al. (1991). Infrastructure and regional economic performance: comment. *New England Economic Review*, (Sep):47–58.

Evans, P. and Karras, G. (1994). Are government activities productive? evidence from a panel of us states. *The Review of economics and statistics*, pages 1–11.

Gu, J. and Volgushev, S. (2019). Panel data quantile regression with grouped fixed effects. *Journal of Econometrics*, 213(1):68–91.

Holtz-Eakin, D. et al. (1994). Public-sector capital and the productivity puzzle. *The Review of Economics and Statistics*, 76(1):12–21.

Koenker, R. and Bassett Jr, G. (1978). Regression quantiles. *Econometrica: journal of the Econometric Society*, pages 33–50.

Koenker, R. and Hallock, K. F. (2001). Quantile regression. *Journal of economic perspectives*, 15(4):143–156.

Lee, E. R., Noh, H., and Park, B. U. (2014). Model selection via bayesian information criterion for quantile regression models. *Journal of the American Statistical Association*, 109(505):216–229.

Leng, X., Chen, H., and Wang, W. (2021). Multi-dimensional latent group structures with heterogeneous distributions. *Journal of Econometrics*.

Lin, C.-C. and Ng, S. (2012). Estimation of panel data models with parameter heterogeneity when group membership is unknown. *Journal of Econometric Methods*, 1(1):42–55.

Liu, R., Shang, Z., Zhang, Y., and Zhou, Q. (2020). Identification and estimation in panel models with overspecified number of groups. *Journal of econometrics*, 215(2):574–590.

Lu, X. and Su, L. (2017). Determining the number of groups in latent panel structures with an application to income and democracy. *Quantitative Economics*, 8(3):729–760.

Munnell, A. H. et al. (1990). Why has productivity growth declined? productivity and public investment. *New England economic review*, (Jan):3–22.

Nyamekye, I., Fiankor, D.-D. D., and Ntoni, J. O. (2016). Effect of human capital on maize productivity in ghana: A quantile regression approach. *International Journal of Food and Agricultural Economics (IJFAEC)*, 4(1128-2016-91989):125–135.

Su, L. and Ju, G. (2018). Identifying latent grouped patterns in panel data models with interactive fixed effects. *Journal of Econometrics*, 206(2):554–573.

Su, L., Shi, Z., and Phillips, P. C. (2016). Identifying latent structures in panel data. *Econometrica*, 84(6):2215–2264.

Von Luxburg, U. (2007). A tutorial on spectral clustering. *Statistics and computing*, 17(4):395–416.

Wang, J. (2010). Consistent selection of the number of clusters via crossvalidation. *Biometrika*, 97(4):893–904.

Yu, L., Gu, J., and Volgushev, S. (2022). Group structure estimation for panel data–a general approach. *arXiv preprint arXiv:2201.01793*.

Zhang, Y., Wang, H. J., and Zhu, Z. (2019). Quantile-regression-based clustering for panel data. *Journal of Econometrics*, 213(1):54–67.