Panel Data Models with Fixed Effects

Yuxin Wang

March 29, 2021

1 Introduction

Consider the following panel data model

$$Y_{it} = X_{it}'\beta + \lambda_i' F_t + \epsilon_{it} \qquad i \in \{1, \dots, N\}, t \in \{1 \dots, T\}, \tag{1}$$

where $\beta \in \mathbb{R}^p$, and $Y_{it} : \Omega \to \mathbb{R}$, $X_{it} : \Omega \to \mathbb{R}^p$, $\lambda_i : \Omega \to \mathbb{R}^r$, $F_t : \Omega \to \mathbb{R}^r$, $\epsilon_{it} : \Omega \to \mathbb{R}$ are all random variables defined on a probability space Ω. In practice, Y_{it} and X_{it} are observed, the coefficient β, the factor loadings λ_i and factors F_t are unknown, and ϵ_{it} is the error term. We call the product $\lambda_i' F_t$ *fixed effects*, and the model 1 *interactive fixed effects model with r factors*.

The multiplicative structure of fixed effects contains the traditional time invariant fixed effects or additive fixed effects as special cases. Indeed, we obtain

$$Y_{it} = X_{it}'\beta + \alpha_i + \epsilon_{it}, \tag{2}$$

if we set $r = 1$, $\lambda_i = \alpha_i$ and $F_t = 1$. And the additive fixed effects model

$$Y_{it} = X_{it}'\beta + \alpha_i + f_t + \epsilon_{it}, \tag{3}$$

is just the case where

$$r = 2, \quad \lambda_i = \begin{pmatrix} \alpha_i \\ 1 \end{pmatrix}, \text{ and } F_t = \begin{pmatrix} 1 \\ f_t \end{pmatrix}.$$

The estimation of β in (2) and (3) are well-documented in standard textbooks (e.g. Greene (2003)). The basic idea is that we can demean the terms in both sides of the equation so that we obtain the transformed data set which can be estimated consistently using OLS. In the more general case (1), demeaning becomes difficult. Holtz-Eakin et al. (1988) proposed a quasi-differencing method to overcome this difficulty. Pesaran (2006) proposed the common correlated effects (CCE) approach that takes weighted cross sectional averages of regressors to approximate the unobserved factors. The CCE estimator is shown to be consistent in Su and Jin (2012) and Westerlund et al. (2019) for any fixed T as $N \to \infty$.

Alternatively, Bai (2009) proposed the least squared estimator for β assuming that the dimension of factors is known. The factors are estimated by principal component analysis based on Stock and Watson (2002). Moon and Weidner (2015) showed that under stronger conditions on the error term, the least squared estimator is always consistent if one chooses sufficiently many factors.

As a motivation of the fixed effects models we look at the following empirical work by Baltagi and Levin (1992). They studied the demand for cigarettes in US states using the following model:

$$\log C_{it} = \alpha + \beta_1 \log C_{i,t-1} + \beta_2 \log P_{it} + \beta_3 \log Y_{it} + \beta_4 \log Pn_{it} + u_{it}.$$

Here $i \in \{1, \dots, 46\}$ is the numbering of the states, $t \in \{1, \dots, 30\}$ stands for the year between 1963-1992. Moreover, C is the sales of cigarettes per capital, P is the price of cigarettes, Y is per capital disposable income, J is the minimum price of cigarettes in the neighboring states. All of them are measured in real term. The authors claim that the disturbance term can be written as

$$u_{it} = \alpha_i + f_t + \epsilon_{it},$$

where α_i is the fixed effect that could for instance represent the presence or size of Indian communities with special treatments of cigarettes. Meanwhile, f_t reflects nationwide policy interventions in different time periods to reduce cigarette consumption. We can even assume that u_{it} has the more general form

$$u_{it} = \lambda_i' F_t + \epsilon_{it}.$$

The methods introduced later can be used to estimate the slope coefficients in this model.

In this paper we present the main ideas of Bai (2009), Bai and Ng (2002), and Moon and Weidner (2015). We provide the algorithms in Bai (2009) and Bai (2002) in Python, replicate some Monte Carlo simulation results of the papers and investigate the simulation with other data generating processes. In the end, we return to the previous empirical example and look at the estimation using the estimator proposed by Bai (2009).

2 Notations and Assumptions

In this paper we use the following notions for the model

$$Y_{it} = X_{it}'\beta + \lambda_i' F_t + \epsilon_{it} \qquad i \in \{1, \ldots, N\}, t \in \{1 \ldots, T\}:$$

$$\underset{(T\times p)}{X_i} := \begin{pmatrix} X_{i1}' \\ \vdots \\ X_{iT}' \end{pmatrix} \qquad \underset{(T\times 1)}{Y_i} := \begin{pmatrix} Y_{i1} \\ \vdots \\ Y_{iT} \end{pmatrix} \qquad i \in \{1 \ldots, N\};$$

$$\underset{(N\times T)}{X^{(k)}} := \begin{pmatrix} X_{11}^{(k)} & \cdots & X_{1T}^{(k)} \\ \vdots & \ddots & \vdots \\ X_{N1}^{(k)} & \cdots & X_{NT}^{(k)}, \end{pmatrix} \qquad k \in \{1 \ldots, p\}$$

And we set

$$\underset{(N\times r)}{\Lambda} := \begin{pmatrix} \lambda_1' \\ \vdots \\ \lambda_N' \end{pmatrix} \qquad \underset{(T\times r)}{F} := \begin{pmatrix} F_1' \\ \vdots \\ F_T' \end{pmatrix}.$$

For any $r \in \mathbb{N}$, the symbol I_r denotes the $r \times r$ identity matrix.

We make the following assumptions:

1. The factor loadings λ_i are independently and identically distributed with uniformly bounded fourth moments. The time series of factors F_t is ergodic with uniformly bounded fourth moments.

2. We have $F'F = TI_r$ and $\Lambda'\Lambda$ is diagonal.

3. The error term ϵ_{it} has mean 0 and has finite eighth moments. Moreover, it is independently and identically distributed across i and t.

4. The regressors X_{it} are independently and identically distributed across i and are ergodic with respect to t. Moreover, they have finite second moments.

As in the standard cross-sectional linear regression model, we also require X_{it} satisfies certain full rank conditions. The assumptions above are very restrictive and can be weakened substantially as described in Bai (2009) and Moon and Weidner (2015). For example, assumption 3 rules out the possibility of a lagged dependent variable as a regressor. In the appendix of Bai (2009) it is shown that this restriction can be removed. We also note that assumption 2 serves as the identification of F_t and λ_i. It is not a binding restriction that affects the estimation of β.

3 Interactive Fixed Effects Models

The estimation of β in the special case (2) with time-invariant fixed effects is well-known. We can take the mean of both sides over time

$$\bar{Y}_i = \bar{X}_i'\beta + \alpha_i + \bar{\epsilon}_i,$$

and subtract the equation above from (2) to obtain

$$\tilde{Y}_{it} = \tilde{X}_{it}'\beta + \tilde{\epsilon}_{it}, \tag{4}$$

where

$$\tilde{Y}_{it} = Y_{it} - \bar{Y}_i, \tilde{X}_{it} = X_{it} - \bar{X}_i, \tilde{\epsilon}_{it} = \epsilon_{it} - \bar{\epsilon}_i.$$

The pooled estimator of β in (4) is called *within estimator*. The estimation of β in (3) is similar. We can transform the equation to

$$\breve{Y}_{it} = \breve{X}_{it}'\beta + \breve{\epsilon}_{it}, \tag{5}$$

where

$$\breve{Y}_{it} = Y_{it} - \frac{1}{T}\sum_{t=1}^{T} Y_{it} - \frac{1}{N}\sum_{i=1}^{N} Y_{it} + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} Y_{it},$$

and $\breve{X}_{it}$ and $\breve{\epsilon}_{it}$ are defined accordingly. Again, we can use the pooled estimator to estimate β in (5). The two within estimators are consistent in the sense that they converge to the true value in probability for any fixed T as $N \to \infty$. In more general cases, transformations along these lines become difficult. Therefore, we will study an M-estimator proposed by Bai (2009) in the rest of this Section.

3.1 Definition and Properties of the Estimator

Let the sum of squared residuals function SSR be

$$\mathrm{SSR}(\beta, F, \Lambda) = \sum_{i=1}^{N}\sum_{t=1}^{T}\left(Y_{it} - X_{it}'\beta - \lambda_i' F_t\right)^2.$$

We define $(\hat{\beta}, \hat{F}, \hat{\Lambda})$ to be the triple such that the sum of squared residuals is minimized:

$$\mathrm{SSR}(\hat{\beta}, \hat{F}, \hat{\Lambda}) = \inf\left\{\mathrm{SSR}(\beta, F, \Lambda) : \frac{F'F}{T} = I_r, \Lambda'\Lambda \text{ diagonal}\right\}. \tag{6}$$

We call $\hat{\beta}$ the *interactive fixed effects estimator* of β. Under the assumptions in Section 2, the estimator $\hat{\beta}$ converges to β in probability as $N \to \infty$ and $T \to \infty$. The number of rows of the estimator $\hat{F}$ increases with T. Nevertheless, it can be shown that

$$\hat{F}(\hat{F}'\hat{F})^{-1}\hat{F} \xrightarrow{p} F(F'F)^{-1}F$$

as $N \to \infty$ and $T \to \infty$ (Bai, 2009).

3.2 Estimation in Practice

Because SSR is differentiable, we could solve the optimization problem (6) by looking for the critical points of SSR, which is analytically and computationally unattractive. Moreover, the estimator $\hat{\beta}$ does not have a closed form. Bai (2009) therefore proposes one of the following iteration procedures:

1. Begin with some guess $\hat{\beta}^{(0)}$. Estimate F_t and λ_i based on $\hat{\beta}^{(0)}$. And update the estimation of β based on $\hat{F}_t$.

2. Begin with some guess $\hat{\beta}^{(0)}$. Estimate F_t and λ_i based on $\hat{\beta}^{(0)}$. And update the estimation of β based on $\hat{F}_t$ and $\hat{\lambda}_i$.

We still need to elaborate some steps. First we demonstrate how to estimate F_t and λ_i if β is given. Assume that β is known, then we can rewrite (1) as

$$Y_{it} - X_{it}'\beta = \lambda_i' F_t + \epsilon_{it},$$

where the left hand side of the equation is known and the right hand side is unknown. We choose $\hat{F}$ and $\hat{\Lambda}$ such that $\text{SSR}(\beta, \hat{F}, \hat{\Lambda})$ is minimised. Stock and Watson (2002) present the derivation of a first order necessary condition for $\hat{F}$ and $\hat{\Lambda}$. Consequently, the estimator $\hat{F}(\beta)$ is a $T \times r$ matrix, whose columns are the r eigenvectors corresponding to the r largest eigenvalues (in absolute form) of the matrix

$$E(\beta) := \left(Y - \sum_{k=1}^{p} \beta_k X^{(k)}\right)\left(Y - \sum_{k=1}^{p} \beta_k X^{(k)}\right)'. \tag{7}$$

In practice one needs to note that $\hat{F}$ should satisfy $\hat{F}'\hat{F}/T = I_r$. Most statistics software provides an orthonormal set of eigenvectors of a matrix. So the resulting matrix should be multiplied by $\sqrt{T}$ to satisfy this condition. Moreover, the first order condition yields the estimator of Λ:

$$\hat{\Lambda}(\beta, \hat{F}) = \frac{1}{T}\left(Y - \sum_{k=1}^{p} \beta_k X^{(k)}\right)' \hat{F}. \tag{8}$$

Next we describe how to estimate β in the model

$$Y_{it} = X_{it}'\beta + \lambda_i' F_t + \epsilon_{it}$$

given that only F_t is known. In this case, we can view F_t as an instrumental variable. The least squared estimator $\hat{\beta}(F)$ for β is

$$\hat{\beta}(F) = \left(\sum_{i=1}^{N} X_i' M_F X_i \right)^{-1} \sum_{i=1}^{N} X_i' M_F Y_i, \tag{9}$$

where M_F is defined to be the projection matrix

$$M_F = I_T - F(F'F)^{-1}F'.$$

Finally, if both λ_i and F_t are known, then we can write (1) as

$$Y_{it} - \lambda_i' F_t = X_{it}'\beta + \epsilon_{it},$$

where the left hand side of the equation is known. In this case, the least squared estimator $\hat{\beta}(F, \Lambda)$ is simply

$$\hat{\beta}(F, \Lambda) = \left(\sum_{i=1}^{N} X_i' X_i \right)^{-1} \sum_{i=1}^{N} X_i'(Y_i - F\lambda_i). \tag{10}$$

We summarize the discussion above as two algorithms.

Algorithm 1

1. Start with some $\hat{\beta}^{(0)} \in \mathbb{R}^p$.

   Suppose we already have $\hat{\beta}^{(0)}, \ldots, \hat{\beta}^{(n)}$.

2. Set $\hat{F}^{(n)} = \hat{F}(\hat{\beta}^{(n)})$ to be the matrix composed of the r eigenvectors corresponding to the r largest eigenvalues of $E(\hat{\beta}^{(n)})$ defined in (7), such that $\hat{F}(\hat{\beta}^{(n)})'\hat{F}(\hat{\beta}^{(n)}) = TI_r$.

3. Set $\hat{\beta}^{(n+1)} = \hat{\beta}(\hat{F}^{(n)})$ according to (9).

4. Repeat 2 and 3 until $|\hat{\beta}^{(n+1)} - \hat{\beta}^{(n)}|$ is sufficiently small.

Algorithm 2

1. Start with some $\hat{\beta}^{(0)} \in \mathbb{R}^p$.

   Suppose we already have $\hat{\beta}^{(0)}, \ldots, \hat{\beta}^{(n)}$.

2. Set $\hat{F}^{(n)} = \hat{F}(\hat{\beta}^{(n)})$ to be the matrix composed of the r eigenvectors corresponding to the r largest eigenvalues of $E(\hat{\beta}^{(n)})$, such that $\hat{F}(\hat{\beta}^{(n)})'\hat{F}(\hat{\beta}^{(n)}) = TI_r$.

3. Set $\hat{\Lambda}^{(n)} = \hat{\Lambda}(\hat{\beta}^{(n)}, \hat{F}^{(n)})$ according to (8).

4. Set $\hat{\beta}^{(n+1)} = \hat{\beta}(\hat{F}^{(n)}, \hat{\Lambda}^{(n)})$ according to (10).

5. Repeat 2 to 4 until $|\hat{\beta}^{(n+1)} - \hat{\beta}^{(n)}|$ is sufficiently small.

The advantage of Algorithm 1 is that we do not put as many restrictions in estimating β in Step 3 as in Step 4 of Algorithm 2. Nevertheless, Algorithm 2 is preferred in practice because the inverse of the matrix

$$\sum_{i=1}^{N} X_i' X_i$$

in (10) does not need to be updated during the iterations. For this reason, we use Algorithm 2 in our Monte Carlo simulations as well.

Due to the interactive term $\lambda_i' F_t$, the objective function SSR is not globally convex in Λ and F. Therefore, both algorithms generate a sequence $(\hat{\beta}^{(n)})_{n \in \mathbb{N}_0}$ converging to the true value only if the starting value $\hat{\beta}^{(0)}$ is properly chosen. A common practice is to use the pooled or within estimation for $\hat{\beta}^{(0)}$.

3.3 Estimating the Parameter with Unknown Number of Factors

We have shown how to estimate β consistently given the a priori knowledge about the structure of the fixed effects. In reality, the number of factors is usually unknown. In this section we are going to investigate the properties of the interactive fixed effects estimator if the number of factors is set incorrectly. And we will study a consistent estimator of the number of factors. Based on the analysis above we see how to estimate β if the number of factors r is unknown.

3.3.1 Estimation with Incorrect Number of Factors

The definition of interactive fixed effects estimator in (6) is dependent on the dimension r of F_t. To make the discussion clear we let $\hat{\beta}(r)$ denote the estimator if the number of factors is assumed to be r. Suppose for now that F_t takes values in $\mathbb{R}^{r_0}$ in the true model, where r_0 is unknown.

In the OLS regression, we can avoid inconsistent estimation of parameters by increasing the number of explanatory variables. A similar phenomenon is observed in Moon and Weidner (2015). They showed that $\hat{\beta}(r)$ is still consistent if $r \geq r_0$ and ϵ_{it} are independently and identically normally distributed. If $r < r_0$, then $\hat{\beta}(r)$ is not consistent.

The assumption that the error terms ϵ_{it} is independently and identically normally distributed is strong. The Monte Carlo simulations in Moon and Weidner (2015) show that the result seems to hold in spite of deviations from this assumption.

We note that choosing a very large r is not a Pareto improvement of lack of knowledge of r_0, because the convergence rate $\hat{\beta}^r$ can be shown to be less than $\hat{\beta}^{r_0}$ in some situation in Moon and Weidner (2015) if $r > r_0$. This is usually the consequence of the overfitting problem as in the OLS regression with too many regressors. And this motivates the need to estimate r_0 in the following.

3.3.2 Estimation of the Number of Factors

The parameter r_0 can be estimated consistently if β and the upper bound $r_{\max}$ of r_0 are known. Assume that β is given in the interactive fixed effects model (1). We set

$$W_{ij} := Y_{ij} - X'_{ij}\beta = \lambda'_i F_t + \epsilon_{it}$$

to be the residual. We have seen in Section 3.1 how to estimate $\lambda_i^{(r)}$ and $F_t^{(r)}$ if we assume λ_i and F_t take value in $\mathbb{R}^r$. After obtaining $\lambda_i^{(r)}$ and $F_t^{(r)}$, we set

$$\hat{\epsilon}_{it}^{(r)} = W_{ij} - \lambda_i^{(r)\prime} F_t^{(r)}$$

and

$$V(r) = \sum_{i=1}^{N} \sum_{t=1}^{T} |\hat{\epsilon}_{it}^{(r)}|^2.$$

Furthermore, we define

$$PC(r) = V(r) + rg(N,T)$$

and

$$IC(r) = \log V(r) + rg(N,T).$$

where g is a positive function satisfying $g(N,T) \to 0$ and $\min\{N,T\}g(N,T) \to \infty$ as $N \to \infty$ and $T \to \infty$. The function g is called a penalty function. Intuitively, we wish to choose r such that $V(r)$ is as small as possible. Here we estimate r_0 by the minimizer $\hat{r}$ of PC or IC on $\{1, \ldots, r_{\max}\}$. The term $rg(N,T)$ avoids the choice of too large r, hence the overfitting problem. Bai (2002) showed that the estimator $\hat{r}$ is consistent as $N \to \infty$ and $T \to \infty$.

3.3.3 Estimation of β with unknown number of factors

The previous discussion enables us to estimate β *if we have a sufficiently large data set*: We can first estimate β assuming a large r, which yields a consistent estimator $\hat{\beta}(r)$. Using $\hat{\beta}(r)$ we can estimate r_0 by the estimator $\hat{r}$ consistently. Finally, we can estimate $\hat{\beta}(\hat{r})$ again. This method does not work if we only have a small sample size. The reason is that $\hat{\beta}^r$ is not an efficient estimator for $r \geq r_0$. If we begin with a large r, it is likely that the estimation is still far from the true value. We propose that for small data sets we can estimate β using various r. We know that if $r \leq r_0$, then the estimation is inconsistent, hence is not closed to the true value. If r is much larger than r_0, the estimation will also not be closed to the true value as well. So there should be a domain for r, where the estimation is around the true value. And we choose the estimation corresponding to the smallest r in this "stability domain". We will give a more explicit explanation in Section 4 with concrete examples.

4 Monte Carlo Simulations and an Empirical Example

We assess the performance of the estimator by Monte Carlo simulations, and apply the method described in Section 3 to real data that we introduced in Section 1. The number of simulations is set to be 1000. We take advantage of MPI to accelerate simulations.

4.1 Estimation of Slope Coefficients

We start with a time-invariant fixed effects model

$$Y_{it} = \beta_1 X_{it,1} + \beta_2 X_{it,2} + \alpha_i + \epsilon_{it}, \tag{11}$$

where $(\beta_1,\ \beta_2) = (1,\ 3)$. The regressors are generated according to $X_{it,j} = 3 + 2\alpha_i + \eta_{it,j}$, with

$$\eta_{it,j} \overset{\text{i.i.d}}{\sim} N(0,1), \qquad j \in \{1,2\},$$

$$\alpha_i \overset{\text{i.i.d}}{\sim} N(0,1),$$

$$\epsilon_{it} \overset{\text{i.i.d}}{\sim} N(0,4).$$

Table 1 compares how the within estimator and the interactive-effects estimator perform in (11). Both estimators are consistent but the within-group estimator is more efficient than the interactive-effects estimator.

In Figure 1, we visualize the results of estimation for β_1 and β_2. The red circles in the plot stand for the interactive-effects estimator, the blue triangles are the within estimator. As the sample size increases from 10 to 100, the points become increasingly concentrated around the true value. The within-group estimator has a lower variance and is closer to the true value than the interactive-effects estimator in each sample size.

In Figure 2, the kernel density curves of both estimators become wider as the sample size increases, showing a higher probability that estimations will take on the true value. Meanwhile, the average median value in the box plot are closer to the true value as well.

In Figure 3, the within estimator has lower root-mean-square error (RMSE) than the interactive-effects estimator. Here we only show the result for β_1. As sample size increases, the RMSE of both estimators are smaller.

The interactive-effects estimator does not work well in small sample, but it is effective under large N and large T. The bias is decreasing with N and T, as shown in the theory and confirmed in the simulation.

By adding a time fixed effects ξ_t to (11), we get an additive fixed effects model

$$Y_{it} = \beta_1 X_{it,1} + \beta_2 X_{it,2} + \alpha_i + \xi_t + \epsilon_{it}, \tag{12}$$

where $(\beta_1,\, \beta_2) = (1,\, 3)$. Two fixed effects satisfy

$$\alpha_i, \xi_t \overset{\text{i.i.d}}{\sim} N(0,1).$$

Both of them are correlated with the two regressors

$$X_{it,j} = 3 + 2\alpha_i + 2\xi_t + \eta_{it,j},$$

with

$$\eta_{it,j} \overset{\text{i.i.d}}{\sim} N(0,1) \qquad j \in \{1,2\}.$$

The error term satisfies

$$\epsilon_{it} \overset{\text{i.i.d}}{\sim} N(0,4).$$

The simulation result is shown in Table 2 and Figure 4-6. The results are the same as in (11). Both estimators are consistent but the within-group estimator is more efficient than the interactive-effects estimator.

Ideally the estimation results of the additive fixed effects model should not depend on the starting values. A popular choice of starting value is the pooled estimator, but it is not always optimal. Table 5 shows that in this model the estimation using interactive effects estimator can be sensitive to the starting values. Two-way estimator works better as starting values than pooled estimator, with its estimation closer to the true value, which is 1 and 3 in (12).

Next we consider a more general setting

$$Y_{it} = \beta_1 X_{it,1} + \beta_2 X_{it,2} + \mu + \lambda_i' F_t + \epsilon_{it}, \tag{13}$$

where $(\beta_1,\, \beta_2,\, \mu) = (1,\, 3,\, 5)$,

$$\lambda_i = \begin{pmatrix} \lambda_{i1} \\ \lambda_{i2} \end{pmatrix} \overset{\text{i.i.d}}{\sim} N(0, I_2),$$

$$F_t = \begin{pmatrix} F_{t1} \\ F_{t2} \end{pmatrix} \overset{\text{i.i.d}}{\sim} N(0, I_2),$$

The regressors are generated according to

$$X_{it,j} = 1 + \lambda_{i1}F_{t1} + \lambda_{i2}F_{t2} + \lambda_{i1} + \lambda_{i2} + F_{t1} + F_{t2} + \eta_{it,j},$$

with

$$\eta_{it,j} \overset{\text{i.i.d}}{\sim} N(0,1) \qquad j \in \{1,2\}.$$

The regressors are correlated with λ_i, f_t, and the product $\lambda_i' F_t$. The regression error

$$\epsilon_{it} \overset{\text{i.i.d}}{\sim} N(0,4).$$

Simulation results are reported in Table 3. In Figures 7-8, we visualize the results and show that the within estimator is inconsistent, while the interactive-effects estimator has consistent estimations for all the coefficients. Figure 9 shows that the RMSE of the interactive-effects estimator becomes smaller for coefficients β_1, β_2, and μ as sample sizes increases.

The results remain similar and are shown in Table 8 if we generate F_{t1} and F_{t2} as independent AR(1) series. That is, we set

$$F_{t,j} = 0.7 \cdot F_{t-1,j} + u_{t,j},$$

with

$$u_{t,j} \overset{\text{i.i.d}}{\sim} N(0,1) \qquad j \in \{1,2\}.$$

The other settings are the same as in (13).

Finally we consider the model

$$Y_{it} = \beta_1 X_{it,1} + \beta_2 X_{it,2} + \mu + x_i \gamma + w_t \delta + \lambda_i' F_t + \epsilon_{it}, \tag{14}$$

where $(\beta_1, \beta_2, \mu, \gamma, \delta) = (1, 3, 5, 2, 4)$,

$$\lambda_i = \begin{pmatrix} \lambda_{i1} \\ \lambda_{i2} \end{pmatrix} \overset{\text{i.i.d}}{\sim} N(0, I_2),$$

$$F_t = \begin{pmatrix} F_{t1} \\ F_{t2} \end{pmatrix} \overset{\text{i.i.d}}{\sim} N(0, I_2).$$

The regressors are generated as before:

$$X_{it,j} = 1 + \lambda_{i1} F_{t1} + \lambda_{i2} F_{t2} + \lambda_{i1} + \lambda_{i2} + F_{t1} + F_{t2} + \eta_{it,j},$$

with

$$\eta_{it,j} \overset{\text{i.i.d}}{\sim} N(0,1) \qquad j \in \{1,2\}.$$

Additionally, we set

$$x_i = \lambda_{i1} + \lambda_{i2} + e_i, \qquad e_i \overset{\text{i.i.d}}{\sim} N(0,1)$$

and

$$w_t = F_{t1} + F_{t2} + \eta_t, \qquad \eta_t \overset{\text{i.i.d}}{\sim} N(0,1),$$

so that x_i is correlated with λ_i and w_t is correlated with f_t.

The simulation results are summarized in Table 4. The table shows that the within-group estimator can only estimate β_1 and β_2 and they are inconsistent. The interactive-effects estimator gives estimation for all the coefficients and shows consistency. In Figures 10-12 we can see the consistency of interactive-effects more clearly. We also did the simulations, where F_{t1} and F_{t2} are AR(1) processes by setting

$$F_{t,j} = 0.7 \cdot F_{t-1,j} + u_{t,j},$$

with

$$u_{t,j} \overset{\text{i.i.d}}{\sim} N(0,1) \qquad j \in \{1,2\}.$$

The other settings are the same as in (14). The estimation of the parameters in this case is still consistent, as shown in Table 9.

4.2 Estimation of the Number of Factors

In the previous models, we know that true factor number r_0 is equal to 2 when we estimate β. In practice, the real number of factors is usually unknown. We estimate the parameters of (14) by using different numbers of factors.

The box plot in Figure 13 shows that the estimator $\hat{\beta}(r)$ is biased and inconsistent if $r < r_0$. However, $\hat{\beta}(r)$ is still consistent if $r \geq r_0$. As shown in Figure 14, with more factors to estimate the RMSE increases as r increases. Therefore, overestimating the factor number is less efficient but still consistent, underestimations of the factor number can be worse with biased results. That is why we usually begin estimations with a relatively large number of factors.

We also estimate the number of factors in (14), by choosing three different penalty functions g_1, g_2, g_3 together with the criteria PC and IC. Table 6 shows that the estimation of $\hat{r}$ is consistent, where the subscript $_{p1}{\sim}_{p3}$ stand for the penalty functions $g_1(N,T) \sim g_3(N,T)$ respectively. Table 6 and Table 7 shows that the true number of factors can be estimated for sufficiently large data sets.

The penalty functions are given by

$$g_1(N,T) = \frac{N+T}{NT} \log \frac{NT}{N+T},$$

$$g_2(N,T) = \frac{N+T}{NT} \log \min\{N,T\},$$

$$g_3(N,T) = \frac{1}{\min\{N,T\}} \log \min\{N,T\}.$$

4.3 Real Data Application

We used the methods that we developed so far to estimate the model

$$\log C_{it} = \alpha + \beta_1 \log C_{i,t-1} + \beta_2 \log P_{it} + \beta_3 \log Y_{it} + \beta_4 \log Pn_{it} + u_{it}.$$

that was mentioned in the introduction of this paper. In the paper Baltagi (1992) considered the structure

$$u_{it} = \alpha_i + f_t + \epsilon_{it}.$$

The original pooled estimation and within estimation of the paper is reported in Table 12. We used the pooled estimation as the starting value, and estimate the coefficients of the model with $R \in \{0, \ldots, 10\}$. Table 10 shows that when $r = 2$, the estimation does not comply with the results of the additive model assumption in the original paper. We observed that when $r \in \{5, 7\}$, the estimation of all coefficients tends to stabilize. This could also be a signal that the true number of factors lies between 5 and 7; the estimations corresponding to $r < 5$ are inconsistent, and small sample size does not guarantee that the estimations corresponding to $r > 7$ converge. We also estimate the number of factors by inserting the estimations from different r. In Table 11, we estimate r by inserting the estimation of the coefficients based on r chosen in the first column. Moreover we set $r_{\max} = 10$. We observe a similar phenomenon as in Moon and Weidner (2015): the most common results are just $r_{\max} = 10$. However, we also notice that many estimations lie between 5 and 7, especially when we focus on the estimation of r between the fifth and the seventh line. This is what the theory expects.

Reference

Bai, J. (2009). Panel data models with interactive fixed effects. Econometrica, 77(4), 1229-1279.

Bai, J., & Ng, S. (2002). Determining the number of factors in approximate factor models. Econometrica, 70(1), 191-221.

Baltagi, B. H., & Levin, D. (1992). Cigarette taxation: Raising revenues and reducing consumption. Structural Change and Economic Dynamics, 3(2), 321-335.

Greene, W. H. (2003). Econometric analysis. Pearson Education.

Holtz-Eakin, D., Newey, W., & Rosen, H. S. (1988). Estimating vector autoregressions with panel data. Econometrica: Journal of the econometric society, 1371-1395.

Moon, H. R., & Weidner, M. (2015). Linear regression for panel with unknown number of factors as interactive fixed effects. Econometrica, 83(4), 1543-1579.

Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. Econometrica, 74(4), 967-1012.

Stock, J. H., & Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. Journal of the American statistical association, 97(460), 1167-1179.

Su, L., & Jin, S. (2012). Sieve estimation of panel data models with cross section dependence. Journal of Econometrics, 169(1), 34-47.

Westerlund, J. (2019). On estimation and inference in heterogeneous panel regressions with interactive effects. Journal of Time Series Analysis, 40(5), 852-857.

Appendix

Table

Table 1: Time Invariant Fixed Effects Model

N	T	Mean $\beta_1 = 1$	SD	Mean $\beta_2 = 3$	SD
Within-Group Estimator					
100	10	0.994	0.070	3.004	0.064
100	20	1.002	0.046	3.000	0.046
100	50	1.001	0.029	3.000	0.028
100	100	1.000	0.020	3.001	0.020
10	100	0.999	0.064	3.002	0.065
20	100	1.003	0.044	3.001	0.043
50	100	0.999	0.028	3.001	0.028
Interactive-Effects Estimator					
100	10	1.019	0.082	3.029	0.082
100	20	1.019	0.059	3.017	0.058
100	50	1.011	0.038	3.011	0.037
100	100	1.006	0.026	3.006	0.026
10	100	1.038	0.103	3.043	0.103
20	100	1.029	0.074	3.027	0.075
50	100	1.012	0.042	3.014	0.041

[a] We report the average value (Mean) and root-mean-square error (SD) of each coefficient over 1000 repetitions in (11).

[b] The true value of β_1 and β_2 are 1 and 3 accordingly.

Table 2: Additive Fixed Effects Model I

N	T	Mean $\beta_1 = 1$	SD	Mean $\beta_2 = 3$	SD
Within Estimator					
100	10	1.002	0.067	3.001	0.066
100	20	0.999	0.046	2.998	0.047
100	50	1.001	0.028	3.000	0.029
100	100	0.999	0.020	3.001	0.020
10	100	1.000	0.070	3.003	0.067
20	100	1.002	0.045	3.001	0.046
50	100	1.000	0.028	3.000	0.028
Interactive-Effects Estimator					
100	10	1.114	0.150	3.113	0.147
100	20	1.100	0.134	3.100	0.135
100	50	1.050	0.090	3.049	0.091
100	100	1.007	0.031	3.009	0.032
10	100	1.117	0.153	3.119	0.153
20	100	1.105	0.136	3.103	0.137
50	100	1.046	0.087	3.046	0.086

[a] We report the average value (Mean) and root-mean-square error (SD) of each coefficient over 1000 repetitions in (12).

[b] The true value of β_1 and β_2 are 1 and 3 accordingly.

Table 3: Interactive Fixed Effects Model I

N	T	Mean $\beta_1 = 1$	SD	Mean $\beta_2 = 3$	SD	Mean $\mu = 5$	SD
Within-Group Estimator							
100	10	1.225	0.234	3.226	0.236	NA	NA
100	20	1.224	0.229	3.220	0.224	NA	NA
100	50	1.221	0.223	3.222	0.224	NA	NA
100	100	1.222	0.224	3.221	0.222	NA	NA
10	100	1.215	0.226	3.216	0.227	NA	NA
20	100	1.218	0.223	3.218	0.223	NA	NA
50	100	1.220	0.223	3.221	0.224	NA	NA
Interactive-Effects Estimator							
100	10	1.073	0.140	3.070	0.137	4.527	1.266
100	20	1.028	0.072	3.028	0.073	4.898	0.264
100	50	1.009	0.036	3.010	0.035	4.974	0.074
100	100	1.005	0.023	3.004	0.023	4.989	0.043
10	100	1.076	0.142	3.073	0.140	4.534	1.149
20	100	1.025	0.074	3.027	0.076	4.893	0.269
50	100	1.007	0.036	3.007	0.034	4.979	0.072

[a] We report the average value (Mean) and root-mean-square error (SD) of each coefficient over 1000 repetitions in (13).
[b] The true value of β_1, β_2 and μ are 1, 3 and 5 accordingly.

Table 4: Interactive Fixed Effects Model with Common Regressors and Time-invariant Regressors I

N	T	Mean $\beta_1 = 1$	SD	Mean $\beta_2 = 3$	SD	Mean $\mu = 5$	SD	Mean $\gamma = 2$	SD	Mean $\delta = 4$	SD
Within-Group Estimator											
100	10	2.091	1.131	4.085	1.125	NA	NA	NA	NA	NA	NA
100	20	2.100	1.117	4.104	1.121	NA	NA	NA	NA	NA	NA
100	50	2.105	1.112	4.102	1.108	NA	NA	NA	NA	NA	NA
100	100	2.104	1.108	4.103	1.107	NA	NA	NA	NA	NA	NA
10	100	2.116	1.130	4.119	1.134	NA	NA	NA	NA	NA	NA
20	100	2.115	1.123	4.112	1.120	NA	NA	NA	NA	NA	NA
50	100	2.111	1.116	4.108	1.113	NA	NA	NA	NA	NA	NA
Interactive-Effects Estimator											
100	10	1.106	0.178	3.107	0.176	4.650	2.958	1.910	0.726	3.951	0.376
100	20	1.032	0.090	3.037	0.093	5.012	1.930	1.979	0.416	3.979	0.330
100	50	1.010	0.039	3.008	0.039	5.059	1.142	1.982	0.395	3.977	0.407
100	100	1.005	0.024	3.005	0.023	5.019	0.992	2.000	0.022	3.989	0.334
10	100	1.108	0.176	3.109	0.176	4.644	2.887	1.915	0.627	3.949	0.599
20	100	1.037	0.090	3.039	0.094	4.950	1.527	1.986	0.198	3.972	0.468
50	100	1.010	0.037	3.009	0.038	5.021	0.862	1.980	0.431	3.994	0.288

[a] We report the average value (Mean) and root-mean-square error (SD) of each coefficient over 1000 repetitions.
[b] In Eq.(14), the true value of β_1, β_2, μ, γ and δ are 1, 3, 5, 2, 4 accordingly.

Table 5: Starting Values in Additive Fixed Effects Model

		pooled			two-way	
N	T	$\beta_1 = 1$	$\beta_2 = 3$		$\beta_1 = 1$	$\beta_2 = 3$
10	100	1.152	3.151		1.109	3.110
20	100	1.155	3.154		1.103	3.103
50	100	1.154	3.157		1.046	3.049
100	10	1.151	3.154		1.108	3.111
100	20	1.157	3.156		1.101	3.100
100	50	1.157	3.154		1.050	3.047
100	100	1.121	3.121		1.008	3.008

[a] We report starting values from pooled estimator and two-way estimator for (12) over over 1000 repetitions.
[b] True value of coefficients are presented in column names.

Table 6: Estimations of Factor Number I

N	T	PC_{p1}	PC_{p2}	PC_{p3}	IC_{p1}	IC_{p2}	IC_{p3}
100	10	8	8	8	8	8	8
100	20	5.098	4.371	6.621	1.812	1.74	1.916
100	50	2	2	2.882	1.997	1.995	2
100	100	2	2	3.647	2	2	2.001
10	100	8	8	8	8	8	8
20	100	5.14	4.381	6.67	1.823	1.752	1.935
50	100	2	2	2.87	2	1.993	2

[a] We report the estimation results of factor number for (14) over 1000 repetitions. Table 6 reports the same sample sizes as Table 1-5.

Table 7: Estimations of Factor Number II

N	T	PC_{p1}	PC_{p2}	PC_{p3}	IC_{p1}	IC_{p2}	IC_{p3}
100	40	2.002	2	3.179	1.994	1.981	1.999
100	60	2	2	2.82	2	1.999	2
200	60	2	2	2	2	2	2
500	60	2	2	2	2	2	2

[a] We report estimation results of factor number for (14) over 1000 repetitions. Table 7 reports a larger sample size than Table 6.

Table 8: Interactive Fixed Effects Model II

N	T	Mean $\beta_1 = 1$	SD	Mean $\beta_2 = 3$	SD	Mean $\mu = 5$	SD
				Within-Group Estimator			
100	10	1.409	0.414	3.405	0.410	NA	NA
100	20	1.425	0.428	3.421	0.423	NA	NA
100	50	1.436	0.437	3.435	0.436	NA	NA
100	100	1.440	0.441	3.440	0.440	NA	NA
10	100	1.432	0.436	3.437	0.441	NA	NA
20	100	1.435	0.437	3.440	0.441	NA	NA
50	100	1.437	0.438	3.439	0.439	NA	NA
				Interactive-Effects Estimator			
100	10	1.092	0.160	3.088	0.158	4.686	2.187
100	20	1.036	0.086	3.030	0.086	4.977	1.247
100	50	1.014	0.037	3.012	0.037	4.975	0.141
100	100	1.010	0.024	3.010	0.023	4.977	0.047
10	100	1.066	0.136	3.070	0.139	4.619	1.262
20	100	1.022	0.069	3.028	0.068	4.912	0.201
50	100	1.011	0.035	3.013	0.035	4.969	0.073

[a] We report the estimation results over 1000 repetitions. We generate Ft1 and Ft2 as independent AR(1) series, the other settings are the same as in (13).

Table 9: Interactive Fixed Effects Model with Common Regressors and Time-invariant Regressors II

N	T	Mean $\beta_1 = 1$	SD	Mean $\beta_2 = 3$	SD	Mean $\mu = 5$	SD	Mean $\gamma = 2$	SD	Mean $\delta = 4$	SD
						Within-Group Estimator					
100	10	1.409	0.414	3.405	0.410	NA	NA	NA	NA	NA	NA
100	20	1.425	0.428	3.421	0.423	NA	NA	NA	NA	NA	NA
100	50	1.436	0.437	3.435	0.436	NA	NA	NA	NA	NA	NA
100	100	1.440	0.441	3.440	0.440	NA	NA	NA	NA	NA	NA
10	100	1.432	0.436	3.437	0.441	NA	NA	NA	NA	NA	NA
20	100	1.435	0.437	3.440	0.441	NA	NA	NA	NA	NA	NA
50	100	1.437	0.438	3.439	0.439	NA	NA	NA	NA	NA	NA
						Interactive-Effects Estimator					
100	10	1.140	0.210	3.136	0.207	4.572	2.052	1.916	0.491	3.880	0.389
100	20	1.067	0.139	3.062	0.137	4.946	1.706	1.951	0.335	3.915	0.404
100	50	1.046	0.115	3.045	0.115	5.336	1.832	1.912	0.528	3.885	0.542
100	100	1.078	0.163	3.078	0.162	5.323	1.762	1.916	0.481	3.790	0.730
10	100	1.155	0.234	3.161	0.241	4.625	2.487	1.868	0.427	3.837	0.517
20	100	1.085	0.174	3.091	0.176	5.228	2.399	1.899	0.469	3.856	0.520
50	100	1.056	0.131	3.057	0.133	5.459	1.959	1.907	0.553	3.825	0.693

[a] We report the estimation results over 1000 repetitions. We generate Ft1 and Ft2 as independent AR(1) series, that is, $F_{t,1} = 0.7 * F_{t-1,1} + u_t$, with $u_t \overset{\text{i.i.d}}{\sim} N(0,1)$. The other settings are the same as in (14).

Table 10: LS Estimation Results. Cigarette Demand Equation 1963–92

Factor number	Intercept	ln $C_{i,t-1}$	ln $P_{i,t}$	ln $Pn_{i,t}$	ln $Y_{i,t}$
R=0	0.278	0.969	-0.090	0.024	-0.031
R=1	0.467	0.905	-0.225	0.075	0.009
R=2	0.887	0.581	-0.357	0.109	0.284
R=3	0.472	0.910	-0.196	0.060	0.003
R=4	2.119	0.363	-0.359	0.115	0.252
R=5	1.563	0.333	-0.343	0.047	0.396
R=6	1.075	0.428	-0.319	0.046	0.396
R=7	1.999	0.335	-0.331	0.042	0.299
R=8	3.491	0.027	-0.334	0.054	0.317
R=9	5.136	-0.159	-0.290	0.091	0.167
R=10	6.005	-0.265	-0.247	0.078	0.094

[a] We report the estimation results over 1000 repetitions.
[b] R $\in \{0, 2, ...10\}$ is the factor number used for estimation.

Table 11: Estimation of Factor Number. Cigarette Demand Equation 1963–92*

R	PC_{p1}	PC_{p2}	PC_{p3}	IC_{p1}	IC_{p2}	IC_{p3}
1	9	9	10	5	2	10
2	10	8	10	6	2	10
3	9	9	10	5	2	10
4	10	9	10	8	4	10
5	10	9	10	7	5	10
6	10	9	10	6	5	10
7	10	9	10	8	4	10
8	10	9	10	10	8	10
9	10	10	10	10	10	10
10	10	10	10	10	10	10

* We set $r_{\max}$=10, and report the estimation results under 1000 repetitions.

Table 12: Estimation of Cigarette Demand Equations in Baltagi (1992)*

	ln $C_{i,t-1}$	ln $P_{i,t}$	ln $Pn_{i,t}$	ln $Y_{i,t}$
OLS	0.969	-0.090	0.024	-0.031
Within	0.833	-0.299	0.034	0.100

* This is a replication of Table 8.1 in Baltagi(1992). We only report the OLS and within estimation results over 1000 repetitions.

Figure

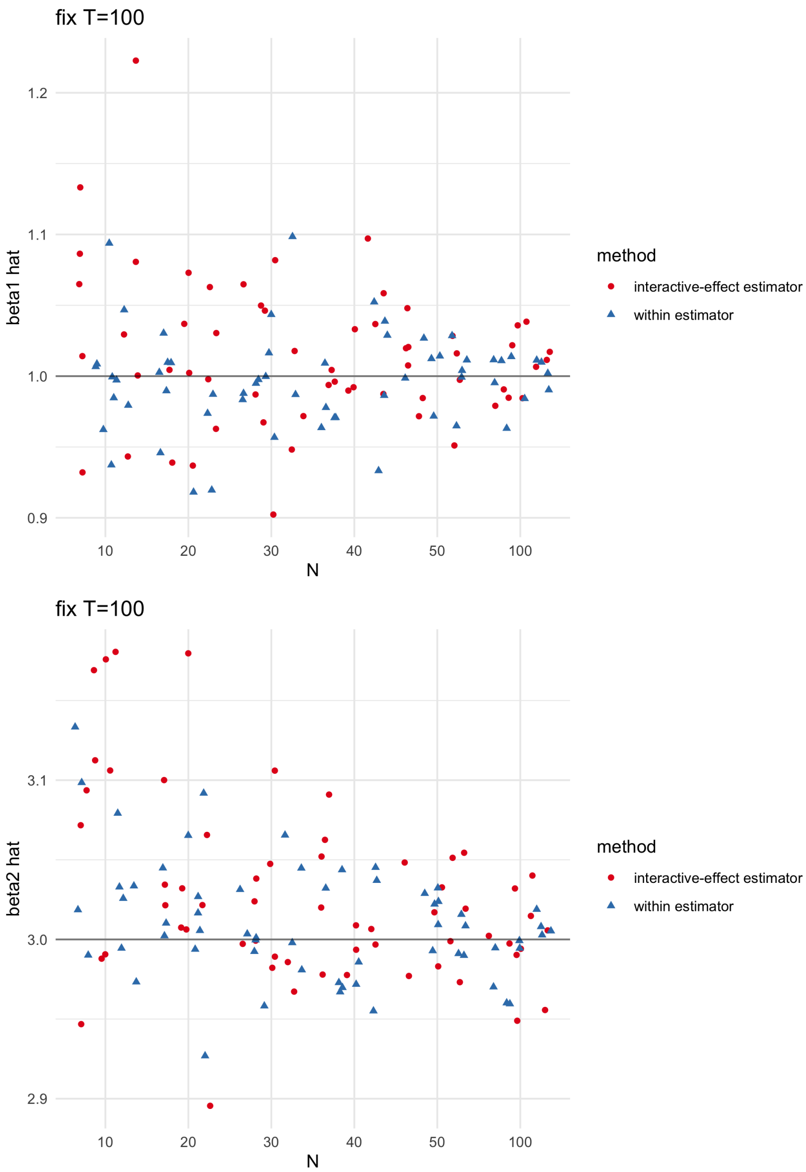


Figure 1: Time Invariant Fixed Effects Model

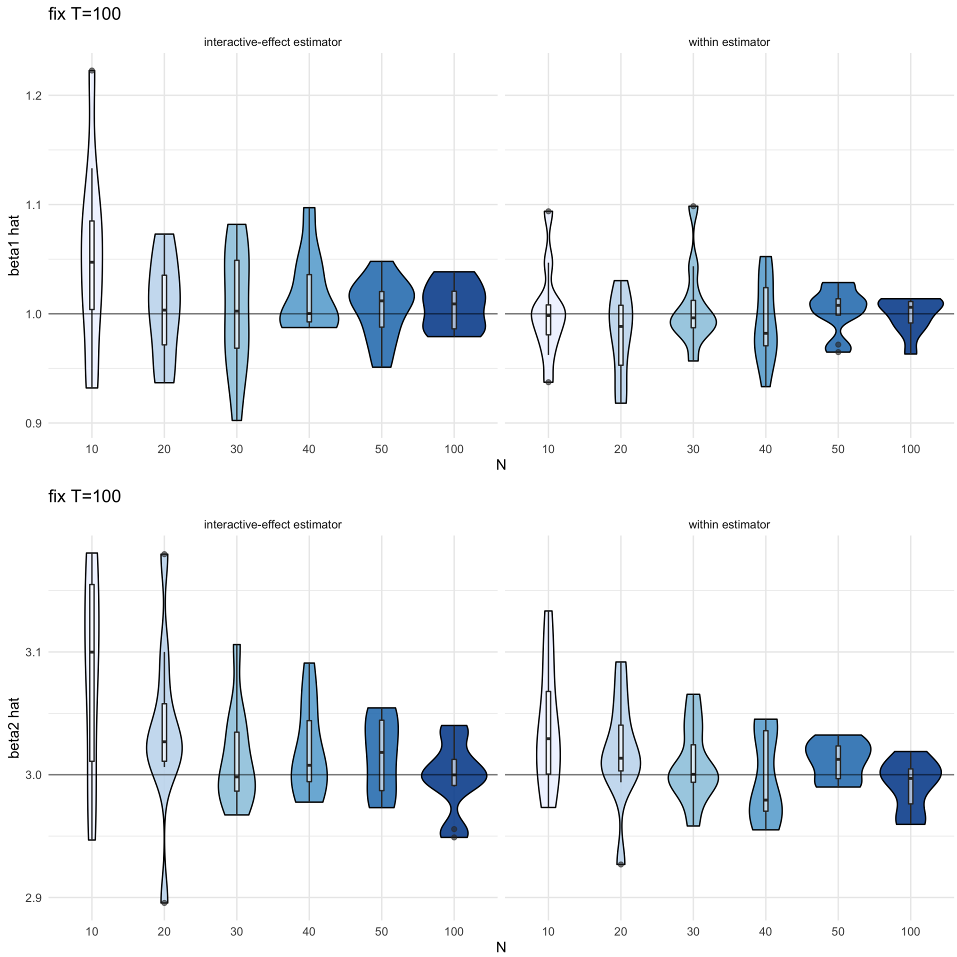
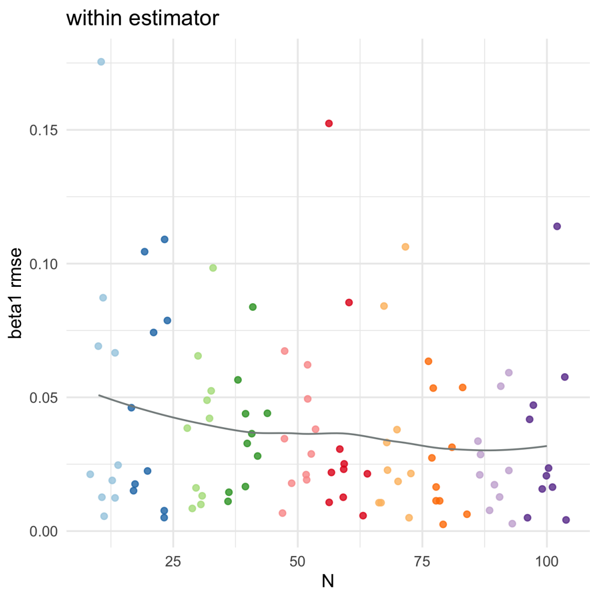


Figure 2: Time Invariant Fixed Effects Model



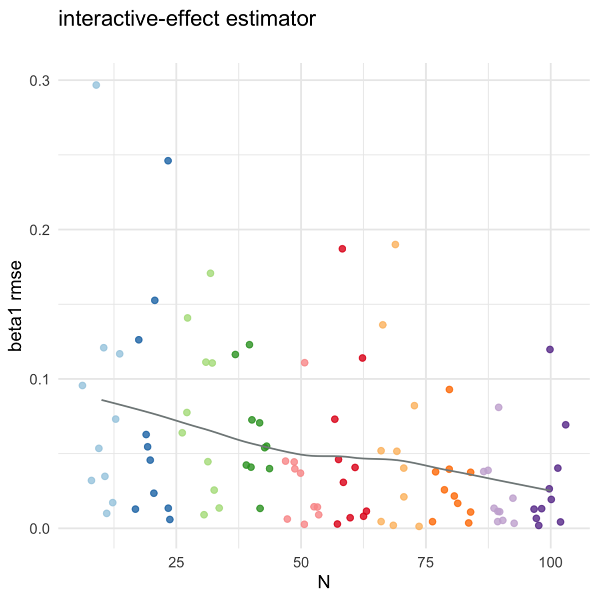


Figure 3: Time Invariant Fixed Effects Model

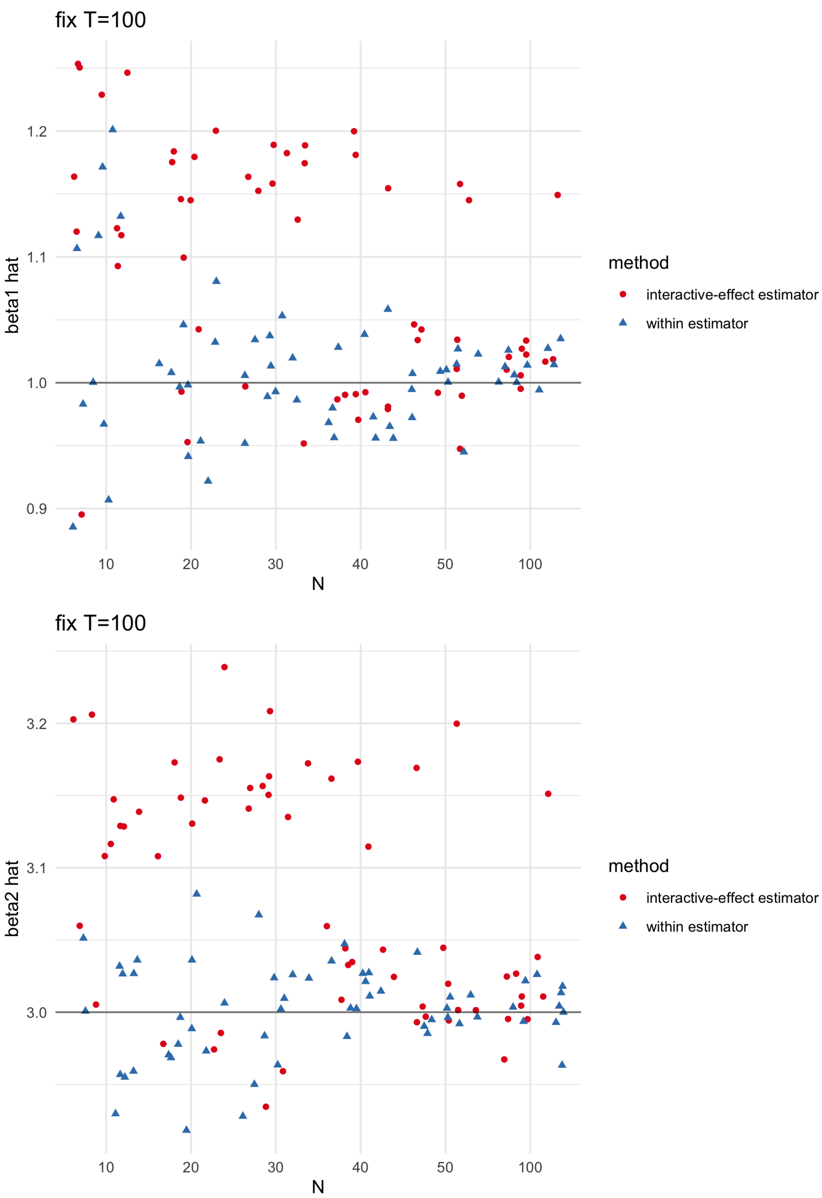


Figure 4: Additive Fixed Effects Model

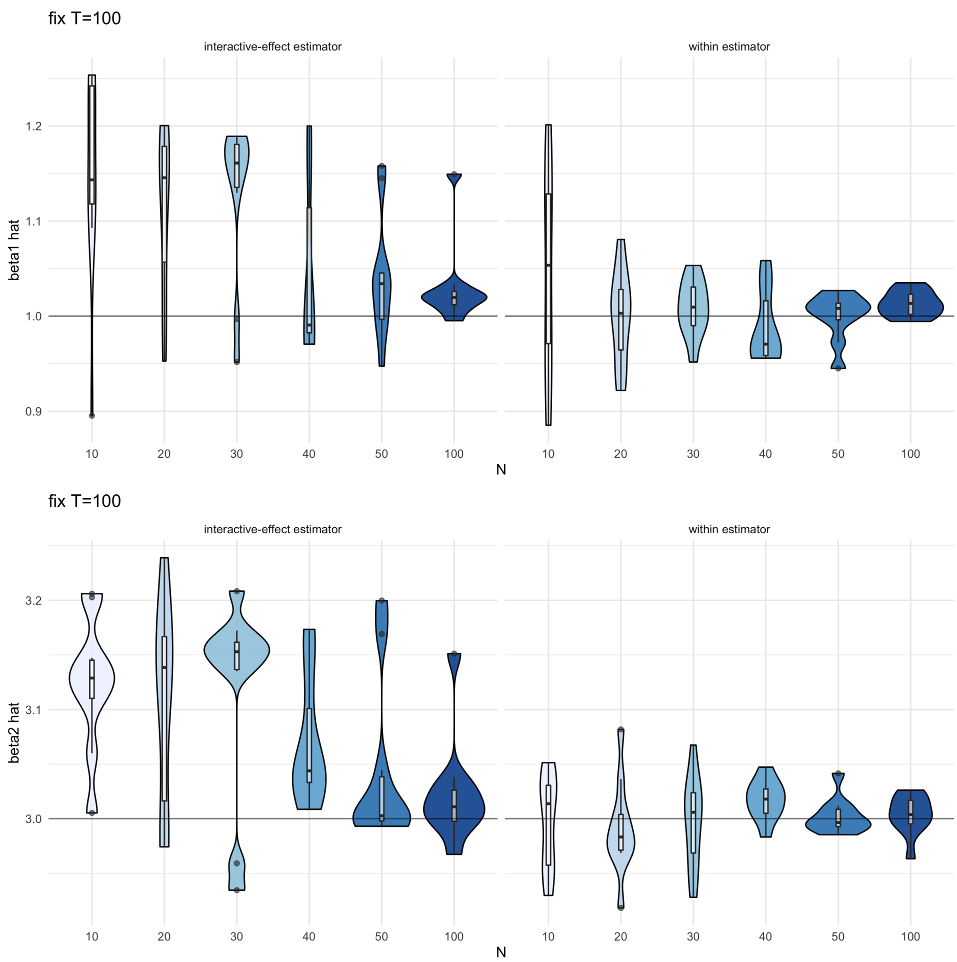


Figure 5: Additive Fixed Effects Model

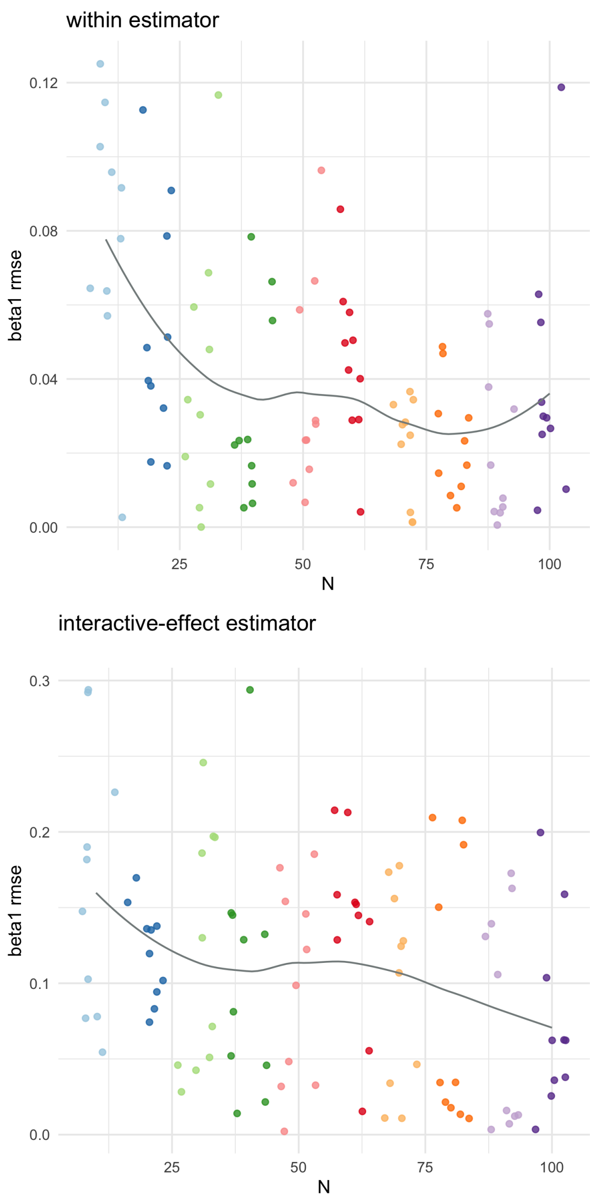


Figure 6: Additive Fixed Effects Model

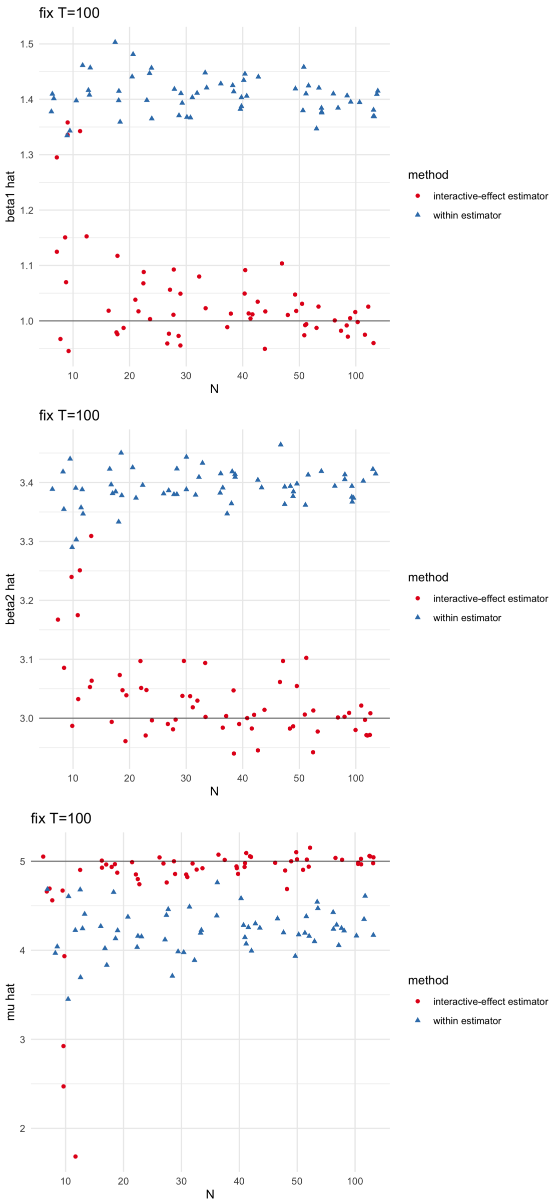


Figure 7: Interactive Fixed Effects Model

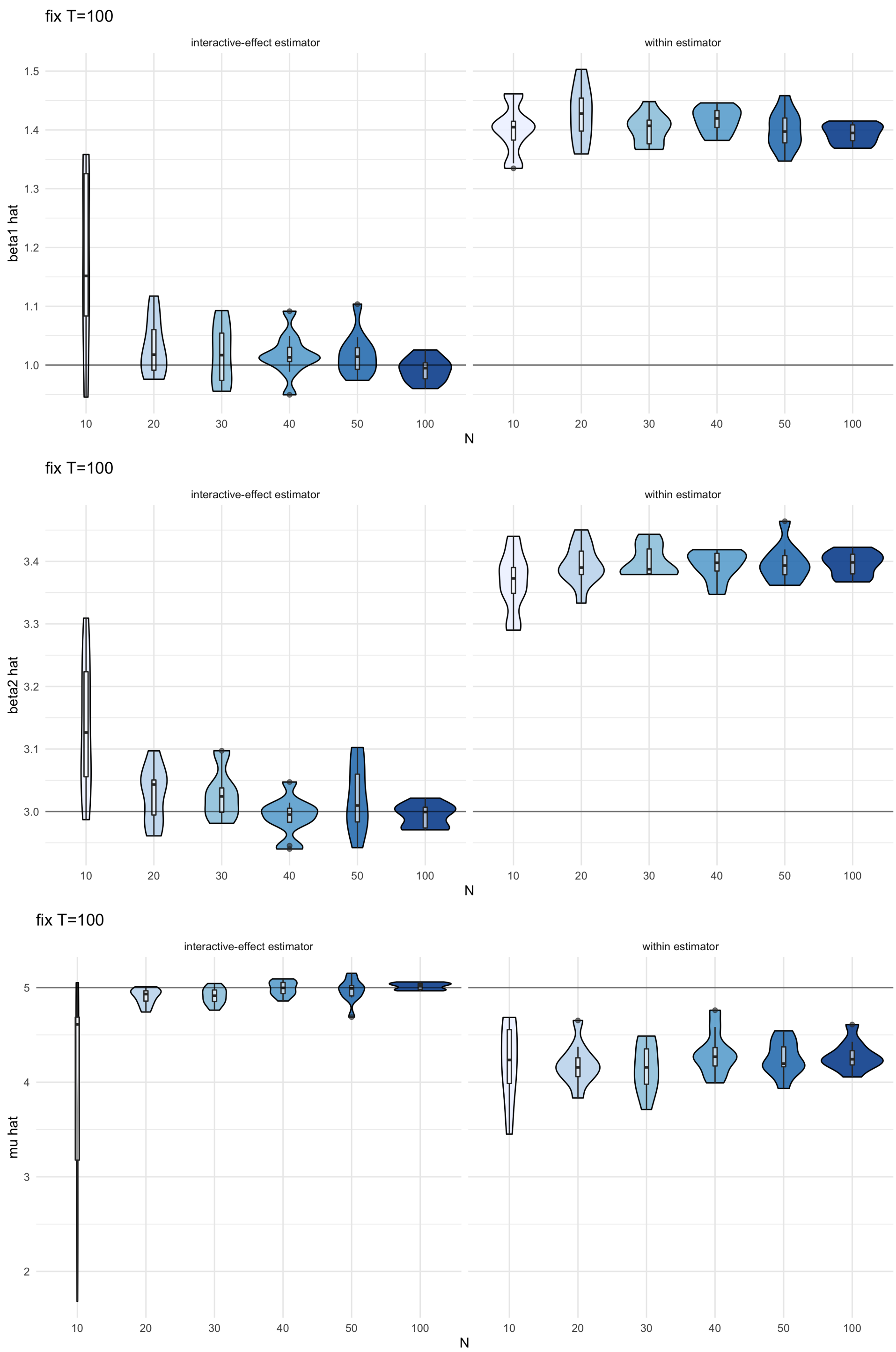


Figure 8: Interactive Fixed Effects Model

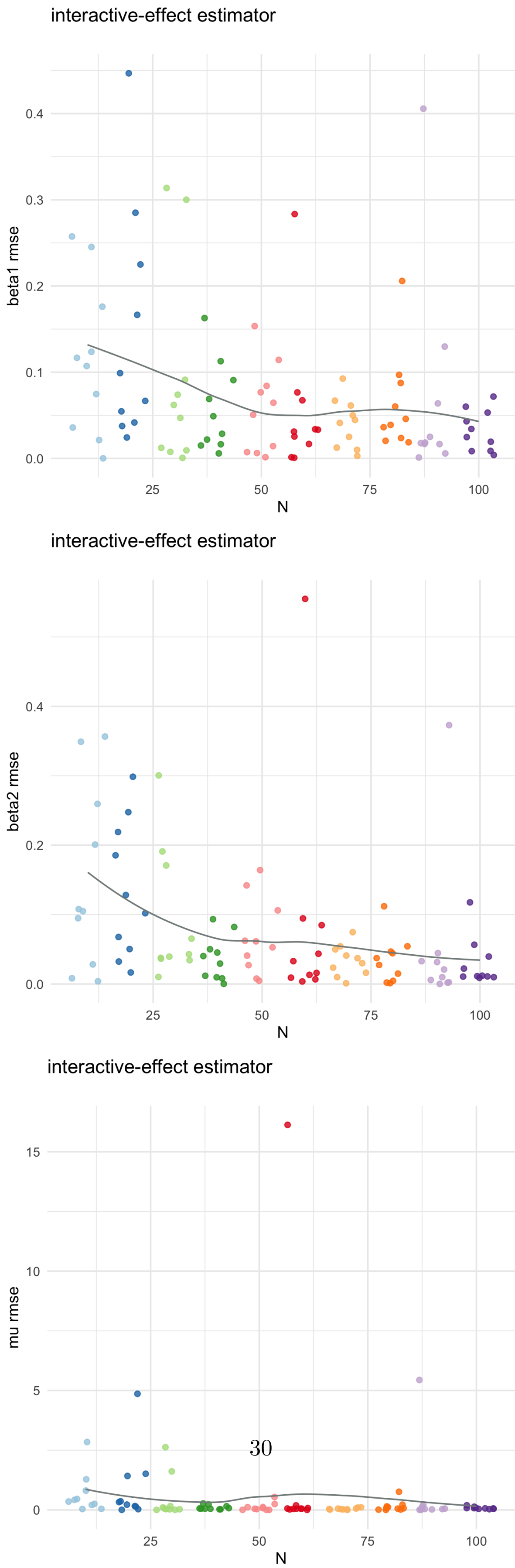


Figure 9: Interactive Fixed Effects Model

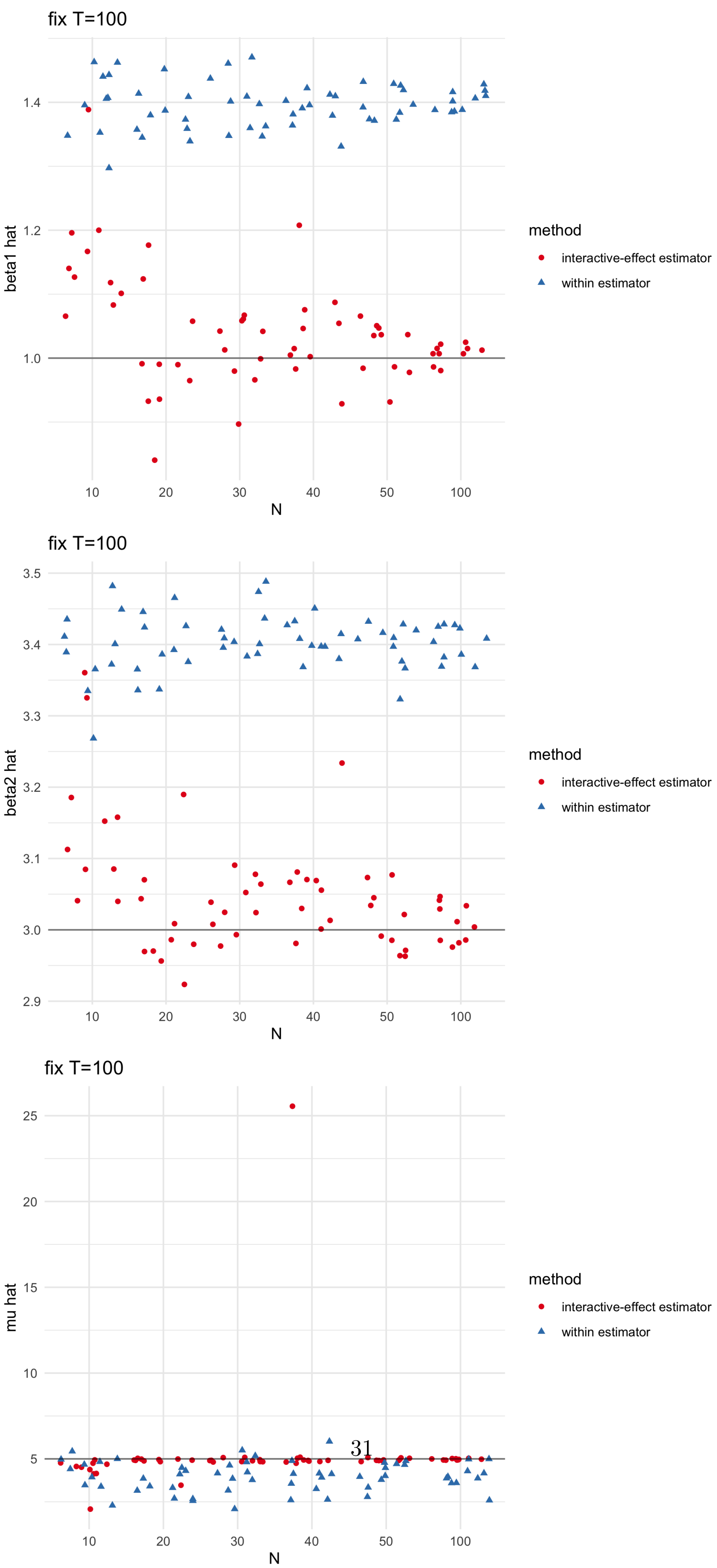


Figure 10: Interactive Fixed Effects Model with Common Regressors and Time-invariant Regressors

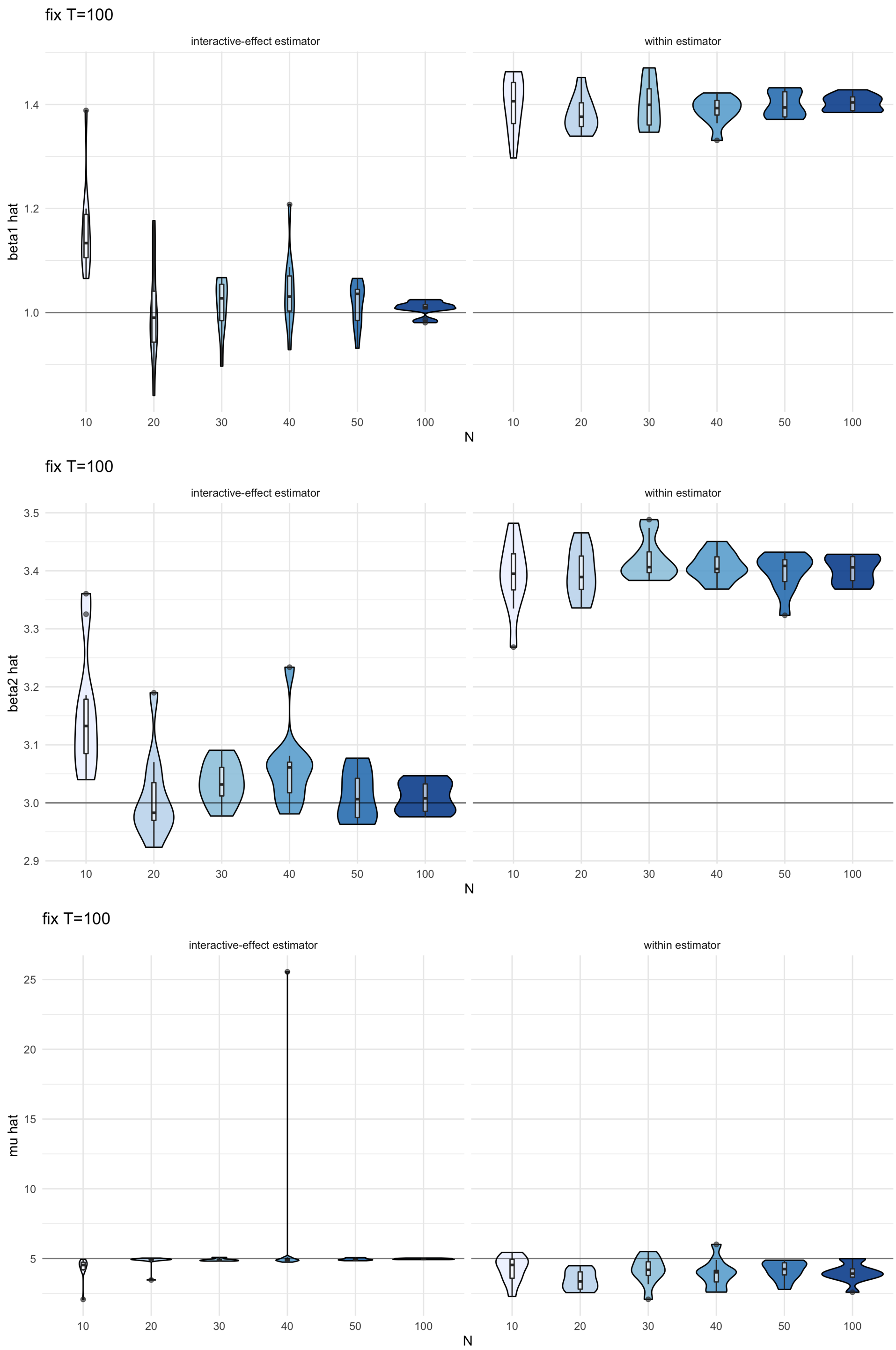


Figure 11: Interactive Fixed Effects Model with Common Regressors and Time-invariant Regressors

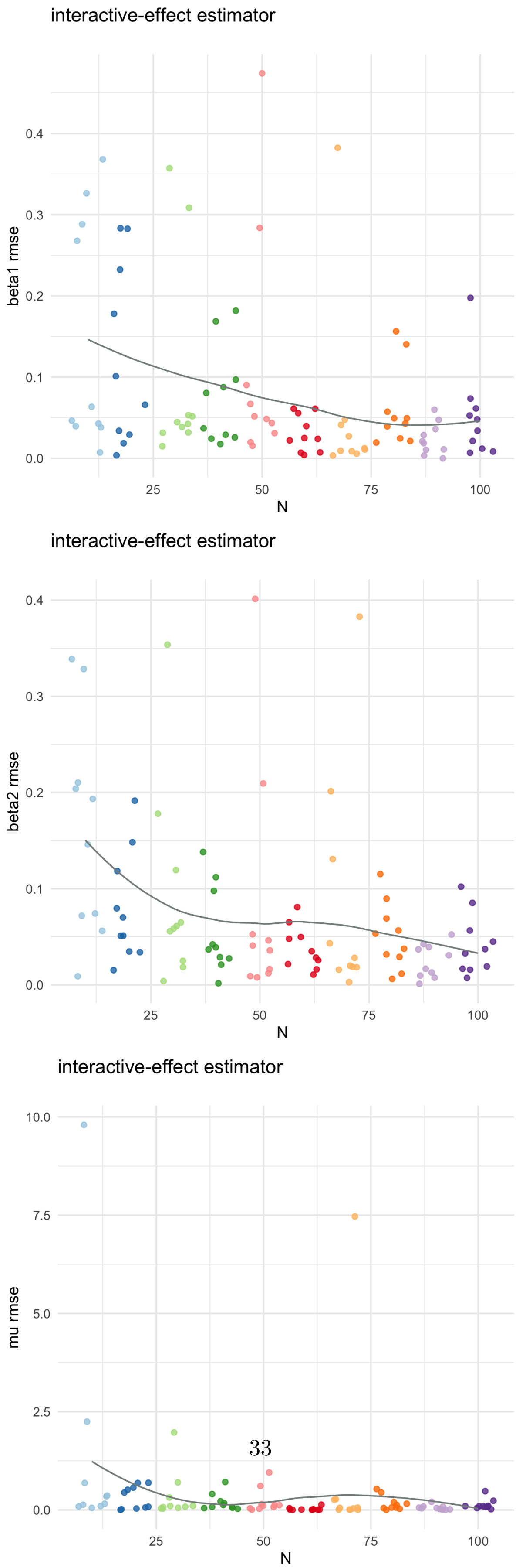


Figure 12: Interactive Fixed Effects Model with Common Regressors and Time-invariant Regressors

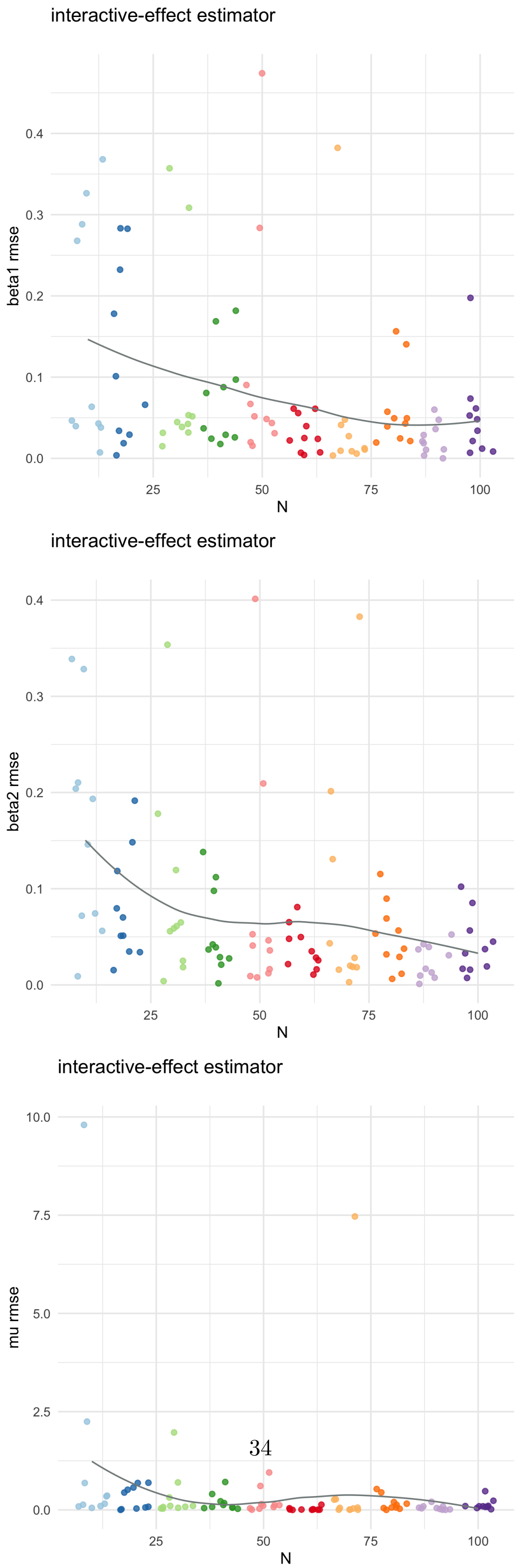


Figure 13: Interactive Fixed Effects Model with Common Regressors and Time-invariant Regressors

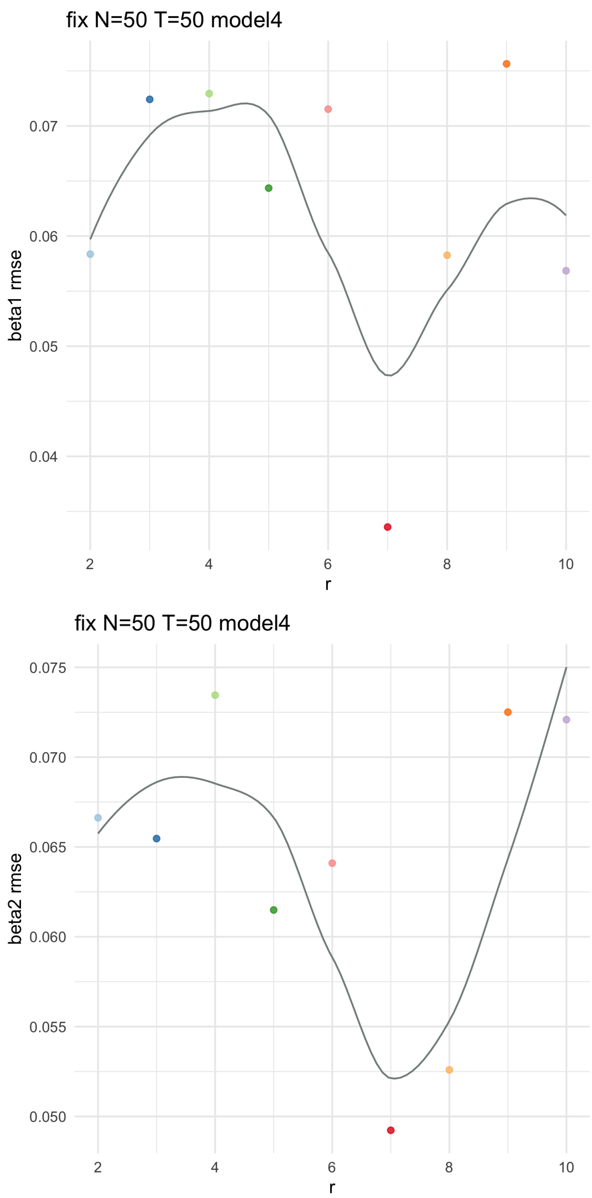

fix N=50 T=50 model4
beta1 rmse
0.07
0.06
0.05
0.04
2
4
6
8
10
r
fix N=50 T=50 model4
beta2 rmse
0.075
0.070
0.065
0.060
0.055
0.050
2
4
6
8
10
r

Figure 14