The London School of Economics and Political Science

# Essays in panel data econometrics with cross-sectional dependence

Lena Körber

A thesis submitted to the Department of Economics of the London School of Economics for the degree of Doctor of Philosophy, London, December 2015.

# Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).



I warrant that this authorization does not, to the best of my belief, infringe the rights of any third party.

I declare that my thesis consists of 60,500 words.

# Statement of conjoint work

I declare that chapters 1 and 2 are jointly co-authored with Oliver Linton and Michael Vogt. This statement is to confirm that I contributed 50 percent to chapter 1 "A Semiparametric Model for Heterogeneous Panel Data with Fixed Effects" and 60 percent to chapter 2 "A Semiparametric Model for Panels of Financial Time Series with an Application to Fragmentation of Trading".

# Abstract

The behavior of economic agents is characterized by interdependencies that arise from common shocks, strategic interactions or spill-over effects. Developing new econometric methodologies for inference in panel data with cross-sectional dependence is a common theme of this thesis. Another theme is econometric models that allow for heterogeneity across individual observations. Each chapter takes a different approach towards modeling and estimating panels with cross-sectional dependence and heterogeneity. In all chapters, the perspective is one where both the time series and the cross-sectional dimension are large.

The first chapter develops a methodology for semiparametric panel data models with heterogeneous nonparametric covariate effects as well as unobserved time and individual-specific effects that may depend on the covariates in an arbitrary way. To model the covariate effects parsimoniously, we impose a dimensionality reducing common component structure on them. In the theoretical part of the chapter, we derive the asymptotic theory of the proposed procedure. In particular, we provide the convergence rates and the asymptotic distribution of our estimators. The asymptotic analysis is complemented by a Monte Carlo experiment that documents the small sample properties of our estimator.

The second chapter investigates the effects of fragmentation in equity markets on the quality of trading outcomes. It uses a unique data set that reports the location and volume of trading on the FTSE 100 and 250 companies from 2008 to 2011 at the weekly frequency. This period coincided with a great deal of turbulence in the UK equity markets which had multiple causes that need to be controlled for. To achieve this, we use the common correlated effects estimator for large heterogeneous panels that approximates the unobserved factors with cross-sectional averages. We extend this estimator to quantile regression to analyze the whole conditional distribution of market quality. We find that both fragmentation in visible order books and dark trading that is offered outside the visible order book lower volatility. But dark trading increases the variability of volatility and trading volumes. Visible fragmentation has the opposite effect on the variability of volatility, in particular at the upper quantiles of the conditional distribution.

The third chapter develops an estimator for heterogeneous panels with discrete outcomes in a setting where the individual units are subject to unobserved common shocks. Like the estimator in chapter 2, the proposed estimator belongs to the class of common correlated effects estimators and it assumes that the unobserved factors are contained in the span of the observed factors and the cross-sectional averages of the regressors. The proposed estimator can be computed by estimating binary response models applied to regression that is augmented with the cross-sectional averages of the individual-specific regressors. The asymptotic properties of this approach are documented as both the time series and the cross-section tend to infinity. In particular, I show that both the estimators of the individual-specific coefficients and the mean group estimator are consistent and asymptotically normal. The small-sample behavior of the mean group estimator is assessed in a Monte Carlo experiment. The methodology is applied to the question of how funding costs in corporate bond markets affect the conditional probability of issuing a corporate bond.

# Acknowledgments

The work contained in this thesis has benefited from the support of many people. I would like to thank Oliver Linton for his encouragement and for many insightful discussions. I very much enjoyed working with him and Michael Vogt on chapters 1 and 2. I am also very indebted to Peter Robinson for his advice and substantive comments on my work. Audiences at seminars and conferences and 5 anonymous referees have provided useful feedback on several parts of the thesis.

The uncountable conversations with my fellow researchers and friends at the LSE and the Bank of England have been an invaluable source of motivation that made my doctoral studies much more enjoyable. Finally, I would like to thank Toni Braun for encouraging me to undertake this PhD and the Institute of Monetary and Economic Studies at the Bank of Japan for hosting me for two inspiring visits in 2010 and 2011.

Financial support from the Economic and Social Research Council and Cusanuswerk is most gratefully acknowledged.

# Disclaimer

A part of this thesis was completed while I was employed at the Bank of England. The views in this thesis are those of the author and do not necessarily reflect the views of the Bank of England, the Monetary Policy Committee, the Financial Policy Committee or the Prudential Regulation Authority.

# Contents

# List of Tables

# List of Figures

# Chapter 1

# A Semiparametric Model for Heterogeneous Panel Data with Fixed Effects[1]

This chapter develops a methodology for semiparametric panel data models in a setting where both the time series and the cross section are large. Such settings are common in finance and other areas of economics. Our model allows for heterogeneous nonparametric covariate effects as well as unobserved time and individual specific effects that may depend on the covariates in an arbitrary way. To model the covariate effects parsimoniously, we impose a dimensionality reducing common component structure on them. In the theoretical part of the chapter, we derive the asymptotic theory of the proposed procedure. In particular, we provide the convergence rates and the asymptotic distribution of our estimators. The estimator is shown to have good small sample properties in a Monte Carlo experiment.

[1]This chapter is written in jointly with Oliver Linton and Michael Vogt and a version of it is published in the *Journal of Econometrics*, Volume 188, Issue 2, October 2015, p. 327-345.

## 1.1 Introduction

In this chapter, we develop estimation methodology for semiparametric panel models in a setting where both the time series and the cross-section dimension are large. Such settings have become increasingly common over the last couple of years. In particular, they are frequently encountered in finance and various areas of economics such as industrial organization or labour economics. Cheng Hsiao has been a pioneer in the development of panel data techniques and his monograph (Hsiao (2003)) contains the main methodological background for our work.

We investigate a regression model which has a nonparametric covariate effect along with individual and time specific fixed effects. The covariate effect is allowed to be heterogeneous across individuals, which is feasible given the long time series we are assuming. To restrict the heterogeneity to be of low dimension, we propose a common component structure on the model. In particular, we assume the individual covariate effects to be composed of a finite number of unknown functions that are the same across individuals but loaded up differently for each cross-sectional unit. The covariate effects are thus modelled as linear combinations of a small number of common functions. The individual and time specific effects of the model are allowed to be related to the covariate in quite a general way. This allows a potential channel for endogeneity, which is important in many applications. We recognize that the endogeneity that is permitted is rather limited, but we remark that this type of restriction is extremely widely exploited in empirical microeconomics, see Angrist and Pischke (2009). A rigorous formulation of the model together with a detailed description of its components is given in Section 1.2. The issue of identifying the various model components is discussed in Section 1.3.

Our model can be regarded as an intermediate case between two extremes. The one extreme is the homogeneous model, where the covariate effect is the same for each cross-sectional unit. This is a very common framework which has been investigated in various parametric and semiparametric studies, see for example Hsiao (2003). In a wide range of applications, it is however rather unrealistic to assume that the covariate effect is the same for all individuals. On the other extreme end, there is the fully flexible model without any restrictions on the covariate effects. One example is the classical SURE model. More recently, Chen et al. (2012) among others have studied a semiparametric version of this very general framework. Even though it is highly flexible, it is however not well suited to some applications. In particular, if the number of individuals is in the hundreds or thousands, the estimation output consists of a huge number of individual functions. This makes the model hardly interpretable. Furthermore, the estimation precision may be very low. Our model lies between these two extremes and allows the user to select the degree of flexibility appropriate for the given application.

Our setting falls in the class of semiparametric panel data models for large cross-section and long time series. Most of the models proposed in the literature for this type of panel data are essentially parametric. Some important papers include Moon and Phillips (1999), Bai and Ng (2002), Bai (2003), Bai (2004), and Pesaran (2006). These authors have addressed a variety of issues including nonstationarity, estimation of unobserved factors, and model selection. Most of the work on semiparametric panel models is in the context of short time series, see for example Kyriazidou (1997). Nonparametric additive models have been considered for instance in Porter (1996). More recent articles include Mammen et al. (2009), Qian and Wang (2012), and Hoderlein et al. (2011).

Only recently, there have been a number of contributions to the non- and semiparametric literature on panels with large cross-section and time series dimension. Linton et al. (2009) consider estimation of a fixed effect time series. Atak et al. (2011) are concerned with seasonality and trends in a panel setting; see also Chen et al. (2013a). Connor et al. (2012) consider a semiparametric additive panel model for stock returns driven by observable covariates and unobservable "factor returns". They allow weak dependence in both time and cross-section direction, but the covariates are not time-varying and there is no individual effect. This model is suited for their application but does not allow a channel for endogeneity. The estimation method is made simpler by the fact that each additive term has a different covariate, whereas the common functions in our model all have the same covariate. Kneip et al. (2012) consider a model similar to ours except that they focus on time as the key nonparametric covariate. Moreover, they do not allow individual effects to be related to included covariates, that is, there is no endogeneity in their model.

Our method to estimate the common functions and the parameter vectors which constitute the individual covariate effects is introduced in Section 1.4. The asymptotic properties of the estimators are described in Section 1.5. In Subsection 1.5.2, we derive the uniform convergence rates as well as an asymptotic normality result for our estimators of the common functions. Importantly, the estimators can be shown to converge to the true functions at the uniform rate $\sqrt{\log nT/nTh}$ which is based on the pooled number of data points $nT$ with $n$ being the cross-section dimension and $T$ the length of the time series. Intuitively, this fast rate is possible to achieve because the functions are the same for all individuals. This allows us to base our estimation procedure on information from the whole panel rather than on a single time series corresponding to a specific individual. In Subsection 1.5.3, we investigate the asymptotic behaviour of our parameter estimators. In particular, we show that they are asymptotically normal. As will turn out, the parameters are estimated with the same precision as in the case where the common functions are known. In particular, our estimators have the same asymptotic distribution as the oracle estimators constructed under the assumption that the functions are observed. To investigate the small sample performance of our estimation procedures, Section

1.8 conducts a series of simulation experiments. Overall, our procedures work well even for quite small sample sizes.

To keep the arguments and discussion as simple as possible, we derive our estimation procedure as well as the asymptotic results under the simplifying assumption that the number of common functions is known. In Sections 1.6 and 1.7, we explain how to dispense with this assumption. In particular, we provide a simple rule to select the number of unknown common functions. This complements our estimation procedure and makes it ready to apply to real data.

## 1.2 The model

In this section, we provide a detailed description of our model framework. We observe a sample of panel data $\{(Y_{it}, X_{it}) : i = 1, \dots, n, \ t = 1, \dots, T\}$, where $i$ denotes the $i$-th individual and $t$ is the time point of observation. To keep the notation as simple as possible, we assume that both the variables $Y_{it}$ and $X_{it}$ are real-valued and focus on the case of a balanced panel.

The data are assumed to come from the model

$$Y_{it} = \mu_0 + \alpha_i + \gamma_t + m_i(X_{it}) + \varepsilon_{it}, \tag{1.1}$$

where $\mathbb{E}[\varepsilon_{it}|X_{it}] = 0$. Here, $m_i$ are nonparametric functions which capture the covariate effect, $\mu_0$ is the model constant (which may be deterministic or stochastic) and the variables $\varepsilon_{it}$ are idiosyncratic error terms. The expressions $\alpha_i$ and $\gamma_t$ are unobserved individual and time specific effects, respectively, which may depend on the regressors in an arbitrary way, e.g., $\alpha_i = G_i(X_{i1}, \dots, X_{iT}; \eta_i)$ and $\gamma_t = H_t(X_{1t}, \dots, X_{nt}; \delta_t)$ for some deterministic functions $G_i, H_t$ and random errors $\eta_i, \delta_t$ that are independent of the covariates. As usual there is an identification shortfall here, and to identify the components of the model, we assume that $\mathbb{E}[m_i(X_{it})] = 0$ along with $\sum_{i=1}^n \alpha_i = \sum_{t=1}^T \gamma_t = 0$.

As the functions $m_i$ may differ across individuals, the covariate effect in our model is allowed to be heterogeneous. However, rather than allowing the effect to vary completely freely, we impose some structure on it. In particular, we assume the functions $m_i$ to have the common component structure

$$m_i(x) = \sum_{k=1}^K \beta_{ik} \mu_k(x), \tag{1.2}$$

where $\mu = (\mu_1, \dots, \mu_K)^\intercal$ is a vector of nonparametric component functions and $\beta_i = (\beta_{i1}, \dots, \beta_{iK})^\intercal$ are parameter vectors. Like the functions $\mu$ and the coefficient vectors $\beta_i$, the number of components $K$ is unobserved. Identifying the functions $\mu$ together with the coefficients $\beta_i$ in our setting is not completely straightforward and requires some care. We thus devote a separate section to this issue. In particular, we provide a detailed discussion in Section 1.3.

We are primarily interested in the case where the individual-specific loadings $\beta_i$ are allowed to be correlated with the regressors $X_{it}$. If instead $\beta_i$ are assumed to be random variables that are distributed independently of the regressors, the model in (1.1) and (1.2) reduces to a semiparametric panel data model with homogeneous covariate effects and individual and time fixed effects (but heteroskedastic errors).

The model defined by (1.1) and (1.2) takes into account several issues which are important in a panel data context. To start with, it captures nonlinearities and heterogeneity in the covariate effect in a flexible but parsimonious way. Moreover, since $\mathbb{E}[\alpha_i + \gamma_t | \{X_{it}\}] \neq 0$ in general, the unobserved effects $\alpha_i$ and $\gamma_t$ introduce a simultaneity between the covariates and the dependent variable. This allows a certain type of endogeneity. Our model and the estimation techniques we develop may thus be applied to a number of different empirical problems where heterogeneity and endogeneity are potential issues.[2]

The type of endogeneity allowed for by the unobserved effects $\alpha_i$ and $\gamma_t$ is rather limited, but we remark that this type of restriction is extremely widely exploited in empirical microeconomics, see Angrist and Pischke (2009). An alternative approach to dealing with endogeneity is to introduce instrumental variables, but there are advantages and disadvantages with that approach also. Our model has the benefit of simplicity and is in line with the simple approach to identifying empirical effects espoused both in Angrist and Pischke (2009) and Manski (2008), for example. It is a generalization of standard heterogeneous linear regression panel data models that are widely discussed in Hsiao (2003) and is part of a large developing literature on semiparametric panel models including Atak et al. (2011), Chen et al. (2012), Connor et al. (2012), Chen et al. (2013a), and Chen et al. (2013b) that explore different weakenings of these models.

The elements $\theta = \{\mu_0, \alpha_i, \gamma_t : i = 1, \ldots, n,\ t = 1, \ldots, T\}$ play the role of nuisance parameters in our framework. There is a large number of them which is increasing with the sample size. Nevertheless, we have an even larger number of observations, which enable us to estimate consistently all the unknown quantities of interest. We thus do not face the "incidental parameters problem" (Neyman and Scott (1948)) that is of wide concern in other panel data settings; see Hsiao (2003) for some discussion of this issue.

We take a pragmatic approach to estimation based on first eliminating the nuisance parameters. To achieve this we make use of a fixed effect transformation.

[2] We note that a symmetric type of model where the heterogeneity in the covariate effect is driven by time rather than individual (i.e., $m_t(\cdot)$ instead of $m_i(\cdot)$) may be of interest in some cases.

Denote the time, cross-sectional, and global averages by:

$$\overline{Y}_i = \frac{1}{T}\sum_{t=1}^{T} Y_{it}, \qquad \overline{Y}_t = \frac{1}{n}\sum_{i=1}^{n} Y_{it}, \qquad \overline{\overline{Y}} = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} Y_{it},$$

and define $Y_{it}^{\text{fe}} = Y_{it} - \overline{Y}_i - \overline{Y}_t + \overline{\overline{Y}}$. Now note that

$$\begin{aligned}
Y_{it}^{\text{fe}} &= m_i(X_{it}) + \varepsilon_{it} - \frac{1}{T}\sum_{t=1}^{T} m_i(X_{it}) - \frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it} - \frac{1}{n}\sum_{i=1}^{n} m_i(X_{it}) - \frac{1}{n}\sum_{i=1}^{n}\varepsilon_{it} \\
&\qquad + \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} m_i(X_{it}) + \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\varepsilon_{it} \\
&= m_i(X_{it}) + \varepsilon_{it} + O_p(T^{-1/2}) + O_p(n^{-1/2}), \qquad (1.3)
\end{aligned}$$

where we require the sample averages to converge to their population means at standard rates, see also Assumption (A1) in Section 1.5. (1.3) shows that the nuisance parameters $\theta$ can be eliminated by subtracting sample means from the data, although this method introduces some additional small error terms.

An alternative procedure is based on differencing, which is the most common method in linear models, see Angrist and Pischke (2009). Specifically, let $Y_{ijt}^{\text{did}} = (Y_{it} - Y_{it-1}) - (Y_{jt} - Y_{jt-1})$ denote the difference-in-difference transformation. Then we have

$$Y_{ijt}^{\text{did}} = (m_i(X_{it}) - m_i(X_{it-1})) - (m_j(X_{jt}) - m_j(X_{jt-1})) + u_{ijt}, \qquad (1.4)$$

where $u_{ijt} = (\varepsilon_{it} - \varepsilon_{it-1}) - (\varepsilon_{jt} - \varepsilon_{jt-1})$ is a serially dependent error term. This approach also eliminates the nuisance parameters $\theta$, but also not completely without cost. First of all, the right-hand side of (1.4) is an additive regression function of the covariates $X_{it}, X_{it-1}, X_{jt}, X_{jt-1}$. To estimate this function, either higher dimensional smoothing must be employed, see Linton and Nielsen (1995), or iterative smoothing techniques like backfitting, see Mammen et al. (1999). Second, the error term $u_{ijt}$ is dependent across time and cross-section, in particular it has a four term "dyadic" (Fafchamps and Gubert (2007)) structure that needs to be accounted for. Finally, one needs stronger conditional moment restrictions on the original error terms to be able to consistently estimate this model. Specifically, we require $\mathbb{E}[\varepsilon_{it}|X_{it}, X_{it-1}, X_{jt}, X_{jt-1}] = 0$ rather than just the assumption $\mathbb{E}[\varepsilon_{it}|X_{it}] = 0$ that will be needed for the fixed effect method. Henderson et al. (2008) propose this method (with just time differencing) in the homogeneous one way model, i.e., $Y_{it} = \mu_0 + \alpha_i + m(X_{it}) + \varepsilon_{it}$.

## 1.3 Identification

The individual regression functions $m_i$ in our model are identified through the normalizations $\mathbb{E}[m_i(X_{it})] = 0$ along with $\sum_{i=1}^n \alpha_i = \sum_{t=1}^T \gamma_t = 0$. We now describe how to identify the vector of common component functions $\mu = (\mu_1, \dots, \mu_K)^\intercal$ and the parameter vectors $\beta_i = (\beta_{i1}, \dots, \beta_{iK})^\intercal$ which constitute the functions $m_i$. Roughly speaking, the idea is to characterize $\mu$ and the parameter vectors $\beta_i$ via an eigenvalue decomposition of a matrix related to the functions $m_i$. Exploiting the uniqueness properties of this decomposition, we are able to identify $\mu$ and the parameter vectors up to sign. Our strategy is thus very similar to the arguments usually used in factor analysis which can for example be found in Connor and Korajczyk (1988) and Bai (2003).

To lay out our strategy, we denote the vector of individual functions by $m = (m_1, \dots, m_n)^\intercal$ and define $B$ to be a $n \times K$ matrix with the entries $\beta_{ik}$ for $i = 1, \dots, n$ and $k = 1, \dots, K$. With this notation at hand, we can represent the vector of functions $m$ as

$$m = B\mu. \tag{1.5}$$

We now put some slight regularity conditions on $B$ and $\mu$. In particular, the functions $\mu$ are assumed to be orthonormal with respect to a weighting function $w$, i.e., $\int \mu(x)\mu(x)^\intercal w(x)dx = I_K$. Moreover, the coefficient matrix $B$ is supposed to have full rank $K$. These assumptions are rather harmless. In particular, the rank condition on $B$ just makes sure that there is enough variation in the coefficients, i.e., in the linear combinations of the $\mu$-functions across individuals.

The above two assumptions on $\mu$ and $B$ can be replaced by a condition which parallels the set of assumptions usually used in factor analysis. In particular, they are equivalent to the following condition:

(I1) The matrix $B$ is orthonormal, i.e. $B^\intercal B = I_K$, and $\int \mu(x)\mu(x)^\intercal w(x)dx$ is a diagonal matrix with non-zero diagonal entries.

To see this equivalence, assume that we start off with a matrix $B^{(1)}$ of rank $K$ and a vector of common component functions $\mu^{(1)}$ which are orthonormal in the sense specified above. Then consider the symmetric, positive definite $K \times K$ matrix $(B^{(1)})^\intercal B^{(1)} = ODO^\intercal$, where $OO^\intercal = O^\intercal O = I_K$ and $D$ is a diagonal matrix with positive entries. Let

$$B^{(2)} = B^{(1)} O D^{-1/2} \tag{1.6}$$

$$\mu^{(2)}(x) = D^{1/2} O^\intercal \mu^{(1)}(x). \tag{1.7}$$

Then

$$(B^{(2)})^\intercal B^{(2)} = D^{-1/2} O^\intercal (B^{(1)})^\intercal B^{(1)} O D^{-1/2} = I_K$$

and

$$\int \mu^{(2)}(x)\mu^{(2)}(x)^\intercal w(x)dx = D^{1/2}O^\intercal O D^{1/2} = D.$$

Hence, the normalized versions $B^{(2)}$ and $\mu^{(2)}$ satisfy (I1).

Let us now assume that the matrix $B$ and the component functions $\mu$ are normalized according to (I1). In addition, suppose that the functions $\mu$ satisfy the following constraint:

(I2) The diagonal entries of the matrix $\int \mu(x)\mu(x)^\intercal w(x)dx$ are all distinct.

This assumption is needed to ensure that the eigenspaces in the spectral decomposition below are one-dimensional, which in turn makes sure that the eigenvectors of the decomposition are uniquely identified up to sign.

Given (I1) and (I2), the matrix $B$ can be characterized via the "covariance" structure of the functions $m$. In particular, we have that

$$\Omega := \int m(x)m(x)^\intercal w(x)dx = B \int \mu(x)\mu(x)^\intercal w(x)dx \ B^\intercal = BDB^\intercal,$$

where $D$ is a diagonal matrix with the diagonal entries $\int \mu_k^2(x)w(x)dx$ for $k = 1,\ldots,K$. These entries are the non-zero distinct eigenvalues of the matrix $\Omega$. Moreover, the columns of the matrix $B$ are the corresponding orthonormal eigenvectors. This spectral decomposition is unique up to the sign of the eigenvectors, i.e., up to the sign of the columns of the matrix $B$. Thus, the coefficients contained in the matrix $B$ are identified up to sign as well.

Exploiting the fact that the columns of $B$ are orthonormal, we can moreover represent the vector of functions $\mu$ by writing

$$\mu = B^\intercal m.$$

This equation almost surely identifies the functions $\mu$ up to sign: The functions $m_i$ contained in the vector $m$ are identified almost surely by our normalizing assumptions. Moreover, as seen above the columns of the matrix $B$ are identified up to sign. As a result, the functions $\mu$ are almost surely identified up to sign as well.

Rather than working with the system (1.5) of dimension $n$ directly, we transform it into a system of dimension $K$. Let $W = (\omega_{ki})$ be a $K \times n$ weighting matrix of rank $K$. Then we can write $Wm = WB\mu$. Introducing the shorthands $S = WB$ and $g = Wm$, we obtain that

$$g = S\mu. \tag{1.8}$$

Here, $g = (g_1,\ldots,g_K)^\intercal$ are weighted averages of the individual functions $m_i$ given by $g_k = \sum_{i=1}^n \omega_{ki} m_i$. Moreover, the $K \times K$ matrix $S$ contains weighted averages of the model parameters as its elements, in particular $S = (s_{kl})$ with $s_{kl} = \sum_{i=1}^n \omega_{ki}\beta_{il}$

for $k, l = 1, \dots, K$. Note that the vectors $m$ and $g$ as well as the matrices $B$, $W$, and $S$ depend on the cross-section dimension $n$. To keep the notation readable, this dependence is suppressed throughout the chapter.

Premultiplying the $n$-dimensional system (1.5) with the matrix $W$, we form $K$ different weighted averages of the individual functions $m$. We thus replace the system (1.5) which characterizes the individual functions $m$ as linear combinations of the common components $\mu$ by a system which represents weighted averages of these functions as linear combinations of $\mu$. The reason for this is twofold: Firstly, the system (1.8) has a fixed dimension $K$ rather than a growing dimension $n$, which is technically more convenient. Secondly, the functions $g$ being averages of the individual functions $m$, they can be estimated much more precisely than the latter. In particular, $g$ can be estimated with a much faster convergence rate than the individual functions. This will help us to achieve a fast convergence rate for our estimator of $\mu$ as well.

The elements of the system (1.8) can be normalized in an analogous way as those of the system (1.5): To start with, we assume that the matrix $S$ has full rank $K$ and that the functions $\mu$ are orthonormal, i.e. $\int \mu(x)\mu(x)^\intercal w(x)dx = I_K$. By the same arguments as before, this is equivalent to the following assumption:

($\mathrm{I_W}1$) The matrix $S$ is orthonormal, i.e. $S^\intercal S = I_K$, and $\int \mu(x)\mu(x)^\intercal w(x)dx$ is a diagonal matrix with non-zero diagonal entries.

Note that the normalization of the functions $\mu$ in ($\mathrm{I_W}1$) depends on the matrix $S$ and thus on the chosen weighting matrix $W$. This becomes visible from equation (1.7) which shows how the normalized version of $\mu$ is constructed. As before, we additionally suppose that the normalized vector of functions $\mu$ has the following property:

($\mathrm{I_W}2$) The diagonal entries of the matrix $\int \mu(x)\mu(x)^\intercal w(x)dx$ are all distinct.

We finally put a slight regularity condition on the weighting scheme $W$:

($\mathrm{I_W}3$) The weights $\omega_{ki}$ are of the form $\omega_{ki} = v_{ki}/n$ with non-negative parameters $v_{ki} \leq C < \infty$ for some sufficiently large constant $C$. For each $k$, the number $n_k$ of nonzero weights is such that $n_k/n \to c_k$ for some positive constant $c_k$.

The above condition is satisfied by a wide range of weighting schemes, for example by the simple choice

$$W = \begin{pmatrix} \overbrace{\tfrac{1}{n} \cdots \tfrac{1}{n}}^{[n/K] \text{ times}} & & & 0 \\ & \tfrac{1}{n} \cdots \tfrac{1}{n} & & \\ & & \ddots & \\ 0 & & & \tfrac{1}{n} \cdots \tfrac{1}{n} \end{pmatrix}. \tag{1.9}$$

Note that by assuming $n_k/n$ to converge to a positive limit, we just make sure that the averages which result from applying the weighting matrix $W$ are composed of $O(n)$ terms. This allows us to apply asymptotic arguments to them later on.

Given the normalization conditions ($\mathrm{I_W}1$) and ($\mathrm{I_W}2$) together with the assumption on the weights ($\mathrm{I_W}3$), the functions $\mu$ can be represented as follows: As the columns of the matrix $S$ are orthonormal, we can write

$$\mu = S^\intercal g. \tag{1.10}$$

The matrix $S$ in this equation can be characterized by a spectral decomposition of the matrix $\Sigma = \int g(x)g(x)^\intercal w(x)dx$. In particular, it holds that

$$\Sigma = S \int \mu(x)\mu(x)^\intercal w(x)dx \; S^\intercal = SDS^\intercal,$$

where $D = \mathrm{diag}(\lambda_1, \dots, \lambda_K)$ with $\lambda_k = \int \mu_k^2(x)w(x)dx$. The constants $\lambda_1, \dots, \lambda_K$ are the non-zero distinct eigenvalues of $\Sigma$. Moreover, the columns of $S$ are the corresponding orthonormal eigenvectors, denoted by $s_1, \dots, s_K$ in what follows.

In the sequel, we shall assume throughout that the functions $\mu$ and the matrix $S$ are normalized to fulfill ($\mathrm{I_W}1$) and ($\mathrm{I_W}2$). Moreover, we suppose that the matrix $\Sigma$ converges to a full-rank matrix $\Sigma^*$. These seem like reasonable and innocuous assumptions. Finally, note that given the existence of a limit $\Sigma^*$, the matrix $S$ converges to a limit $S^*$ as well. This is due to the fact that the eigenvectors $s_1, \dots, s_K$ depend continuously on the entries of the matrix $\Sigma$.

## 1.4 Estimation

We now describe our procedure to estimate the functions $\mu_1, \dots, \mu_K$ and the coefficient vectors $\beta_i = (\beta_{i1}, \dots, \beta_{iK})^\intercal$ based on kernel methods. Of course, alternative methods can be used, including the iterative algorithms developed in Chen et al. (2013a) or the sieve methods described in Chen (2013). One advantage of our procedures is that they are "in closed form" meaning that one does not have to rely on nonlinear optimization and that they can be computed very fast and accurately even with very large datasets.

For simplicity of exposition, we assume throughout the section that the number $K$ of common components is known. In Sections 1.6 and 1.7, we will dispense with this assumption and provide a procedure to estimate $K$. Our approach splits up into four steps, each of which is described in a separate subsection. To start with, we construct preliminary estimators of the individual regression functions $m_i$. These are used to obtain estimators of the $\mu$-functions and the coefficient vectors $\beta_i$ in a second and third step, respectively. In a final step, we exploit the model structure to obtain improved estimators of the individual regression functions $m_i$.

### 1.4.1 Preliminary estimators of the individual functions

We estimate the individual functions $m_i$ by applying nonparametric kernel techniques to the time series data $\{(Y_{it}^{\text{fe}}, X_{it}) : t = 1, \ldots, T\}$. More specifically, Nadaraya-Watson or local linear smoothers may be used. The Nadaraya-Watson estimator of the function $m_i$ is defined as

$$\widehat{m}_i^{\text{NW}}(x) = \frac{\sum_{t=1}^T K_h(x - X_{it}) Y_{it}^{\text{fe}}}{\sum_{t=1}^T K_h(x - X_{it})},$$

where $h$ is a scalar bandwidth and $K(\cdot)$ denotes a kernel satisfying $\int K(u)du = 1$ and $K_h(\cdot) = h^{-1}K(h^{-1}\, \cdot)$. The local linear estimator of $m_i$ is given by the formula

$$\widehat{m}_i^{\text{LL}}(x) = \frac{\sum_{t=1}^T w_{i,T}(x, X_{it}) Y_{it}^{\text{fe}}}{\sum_{t=1}^T w_{i,T}(x, X_{it})},$$

with

$$w_{i,T}(x, X_{it}) = K_h(x - X_{it}) \Big( S_{i,T,2}(x) - \Big( \frac{x - X_{it}}{h} \Big) S_{i,T,1}(x) \Big)$$

and

$$S_{i,T,k}(x) = \frac{1}{T} \sum_{t=1}^T K_h(x - X_{it}) \Big( \frac{x - X_{it}}{h} \Big)^k$$

for $k = 1, 2$; see Fan and Gijbels (1996) for a detailed account of the local linear smoothing method. The procedure to estimate the functions $\mu$ and the parameter vectors $\beta_i$ is the same no matter whether we work with Nadaraya-Watson or local linear smoothers. In what follows, we thus use the symbol $\widehat{m}_i$ to denote either the local constant estimator $\widehat{m}_i^{\text{NW}}$ or the local linear smoother $\widehat{m}_i^{\text{LL}}$.

### 1.4.2 Estimating the common component functions $\boldsymbol{\mu}$

We now use the characterization (1.10) of the functions $\mu$ to construct an estimator of them. We proceed as follows:

Step 1: Construct estimators $\widehat{g} = (\widehat{g}_1, \ldots, \widehat{g}_K)^\intercal$ of the functions $g = (g_1, \ldots, g_K)^\intercal$ according to

$$\widehat{g}_k(x) = \sum_{i=1}^n \omega_{ki} \widehat{m}_i(x).$$

Step 2: Estimate the matrix $\Sigma$ by

$$\widehat{\Sigma} = \int \widehat{g}(x)\widehat{g}(x)^\intercal w(x) dx.$$

Step 3: Estimate the eigenvalues and eigenvectors by

$$\widehat{\Sigma} = \widehat{S}\widehat{D}\widehat{S}^\intercal,$$

i.e., by performing an eigenvalue decomposition of the matrix $\widehat{\Sigma}$. Let $\widehat{\lambda}_1, \dots, \widehat{\lambda}_K$ be the eigenvalues of $\widehat{\Sigma}$ (i.e. the diagonal entries of the matrix $\widehat{D}$), and $\widehat{s}_1, \dots, \widehat{s}_K$ the corresponding orthonormal eigenvectors (i.e. the columns of the matrix $\widehat{S}$).

Step 4: Define the estimator of $\mu$ by replacing $S$ and $g$ in (1.10) with their respective estimators, i.e.,

$$\widehat{\mu} = \widehat{S}^\intercal \widehat{g}.$$

### 1.4.3 Estimating the coefficients $\boldsymbol{\beta_i}$

Consider the time series data $\{(Y_{it}, X_{it}) : t = 1, \dots, T\}$ of the $i$-th individual. These are assumed to come from the model

$$Y_{it} = \mu_0 + \alpha_i + \gamma_t + \sum_{k=1}^{K} \beta_{ik} \mu_k(X_{it}) + \varepsilon_{it}$$

for $t = 1, \dots, T$, which is linear in the parameters $\beta_i = (\beta_{i1}, \dots, \beta_{iK})^\intercal$. If the functions $\mu_1, \dots, \mu_K$ were known, the coefficients $\beta_i$ could be estimated by standard least squares methods from the time series data $\{(Y_{it}^{\text{fe}}, X_{it}) : t = 1, \dots, T\}$. In particular, we could use a weighted least squares estimator given by

$$\widetilde{\beta}_i = \Big( \frac{1}{T} \sum_{t=1}^{T} \pi(X_{it}) \mu(X_{it}) \mu(X_{it})^\intercal \Big)^{-1} \frac{1}{T} \sum_{t=1}^{T} \pi(X_{it}) \mu(X_{it}) Y_{it}^{\text{fe}} \tag{1.11}$$

with a weighting function $\pi$. As the functions $\mu$ are not known, we replace them by the estimates $\widehat{\mu}$, thus yielding the estimator

$$\widehat{\beta}_i = \Big( \frac{1}{T} \sum_{t=1}^{T} \pi(X_{it}) \widehat{\mu}(X_{it}) \widehat{\mu}(X_{it})^\intercal \Big)^{-1} \frac{1}{T} \sum_{t=1}^{T} \pi(X_{it}) \widehat{\mu}(X_{it}) Y_{it}^{\text{fe}}. \tag{1.12}$$

### 1.4.4 Re-estimating the functions $\boldsymbol{m_i}$ and iterating the estimation procedure

Exploiting the model structure, we can now define new estimators of the individual functions $m_i$ which have better asymptotic properties than the preliminary estimators $\widehat{m}_i$. Specifically, we let

$$\widehat{m}_i^e(x) = \widehat{\beta}_i^\intercal \widehat{\mu}(x).$$

As we will see later on, the estimators $\widehat{m}_i^e$ have a faster convergence rate than the preliminary smoothers $\widehat{m}_i$.

A possible extension of our estimation procedure is to iterate it. To do so, we first re-estimate the component functions $\mu$ and the parameters $\beta_i$ by using $\widehat{m}_i^e$ instead of the preliminary smoothers $\widehat{m}_i$. This yields updated estimators of $\mu$ and $\beta_i$. In addition, we may update the estimated individual effects whose first round estimates were implicitly given by $\widehat{\alpha}_i = \overline{Y}_i - \overline{\overline{Y}}$, $\widehat{\gamma}_t = \overline{Y}_t - \overline{\overline{Y}}$, and $\widehat{\mu}_0 = \overline{\overline{Y}}$. Specifically, these may be replaced by:

$$\widehat{\alpha}_i^e = \frac{1}{T}\sum_{t=1}^{T}\left\{Y_{it} - \widehat{\mu}_0 - \widehat{m}_i^e(X_{it})\right\}; \quad \widehat{\gamma}_t^e = \frac{1}{n}\sum_{i=1}^{n}\left\{Y_{it} - \widehat{\mu}_0 - \widehat{m}_i^e(X_{it})\right\};$$

$$\widehat{\mu}_0^e = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left\{Y_{it} - \widehat{m}_i^e(X_{it})\right\}.$$

This process can be continued until some convergence criterion is satisfied, which is likely to be achieved in practice quite quickly. Note that we can view this iterative algorithm as a procedure to find the minimum of a least squares objective function along the lines of Connor et al. (2012).

## 1.5 Asymptotics

In what follows, we derive the asymptotic properties of our estimators. To start with, we list the assumptions needed for our analysis. We then present the results on the limiting behaviour of the estimators $\widehat{\mu}$, $\widehat{\beta}_i$, and $\widehat{m}_i^e$. The proofs of our theoretical results can be found in Appendix A.1.

### 1.5.1 Assumptions

We impose the following regularity conditions, which as usual are sufficient but not necessary for our results. The expression $T^a \ll n \ll T^b$ is used to mean that $CT^{a+\delta} \leq n \leq CT^{b-\delta}$ for some positive constant $C$, a small $\delta > 0$ and $0 < a < b$. The symbol $\gg$ is used analogously.

(A1) The data $\{X_{it} : i = 1, \dots, n,\ t = 1, \dots, T\}$ and disturbances $\{\varepsilon_{it} : i = 1, \dots, n,\ t = 1, \dots, T\}$ are independent across $i$. Moreover, they are strictly stationary and strongly mixing (Rosenblatt (1956)) in the time direction. Let $\alpha_i(k)$ for $k = 1, 2, \dots$ be the mixing coefficients of the time series $\{(X_{it}, \varepsilon_{it}), t = 1, \dots, T\}$ of the $i$-th individual. It holds that $\alpha_i(k) \leq \alpha(k)$ for all $i = 1, \dots, n$, where the coefficients $\alpha(k)$ decay exponentially fast to zero as $k \to \infty$.

(A2) The densities $f_i$ of the variables $X_{it}$ exist and have bounded support, $[0,1]$ say. Moreover, they are uniformly bounded away from zero and from above, i.e., $0 < c \leq \min_{1 \leq i \leq n} \inf_{x \in [0,1]} f_i(x)$ as well as $\max_i \sup_x f_i(x) \leq C < \infty$ for some pair of constants $0 < c \leq C < \infty$. Finally, the joint densities $f_{i;l}$ of $(X_{it}, X_{it+l})$ exist and are also uniformly bounded from above.

(A3) The functions $\mu_1, \dots, \mu_K$ are twice continuously differentiable on $[0,1]$. Moreover, the densities $f_i$ are twice continuously differentiable on $[0,1]$ as well with uniformly bounded first and second derivatives $f_i'$ and $f_i''$. Finally, the coefficients $\beta_{ik}$ are bounded by some constant $\overline{\beta} < \infty$, i.e., $|\beta_{ik}| \le \overline{\beta}$ for all $i = 1, \ldots, n$ and $k = 1, \ldots, K$, which ensures that the functions $m_i$ as well as the derivatives $m_i'$ and $m_i''$ are uniformly bounded on $[0, 1]$ as well.

(A4) It holds that $\mathbb{E}[\varepsilon_{it}|X_{it}] = 0$. Moreover, for some $\theta > 5$ and for all $l \in \mathbb{Z}$,

$$\max_{1\le i\le n} \sup_{x\in[0,1]} \mathbb{E}\big[|\varepsilon_{it}|^\theta \big| X_{it} = x\big] \le C < \infty \tag{1.13}$$

$$\max_{1\le i\le n} \sup_{x,x'\in[0,1]} \mathbb{E}\big[|\varepsilon_{it}| \big| X_{it} = x, X_{it+l} = x'\big] \le C < \infty \tag{1.14}$$

$$\max_{1\le i\le n} \sup_{x,x'\in[0,1]} \mathbb{E}\big[|\varepsilon_{it}\varepsilon_{it+l}| \big| X_{it} = x, X_{it+l} = x'\big] \le C < \infty, \tag{1.15}$$

where $C$ is a sufficiently large constant independent of $l$.

(A5) The cross-section dimension $n = n(T)$ depends on $T$ and satisfies $T^{2/3} \ll n \ll T^{3/2}$.

(A6) The bandwidth $h$ is of the order $(nT)^{-(1/5+\delta)}$ for some small $\delta > 0$.

(A7) The kernel $K$ is bounded, symmetric about zero and has compact support ($[-C_1, C_1]$, say). Moreover, it fulfills the Lipschitz condition that there exists a positive constant $L$ with $|K(u)-K(v)| \le L|u-v|$. Let $\mu_2(K) = \int K(\varphi)\varphi^2 d\varphi$ and $\|K\|_2^2 = \int K^2(\varphi) d\varphi$.

Assumption (A1) is very strong for the type of large panel data sets considered in this paper. By restricting the degree of correlation between the regressors $X_{it}$ and the fixed effects $\alpha_i$ and $\gamma_t$, it rules out situations where the regressors are generated by $X_{it} = \mu_X + \alpha_i + \gamma_t + u_{it}$, for example. But it is possible to relax (A1) in several dimensions as we discuss below.

Note that we do not necessarily require exponentially decaying mixing rates as assumed in (A1). These could alternatively be replaced by sufficiently high polynomial rates. We nevertheless make the stronger assumption (A1) to keep the notation and structure of the proofs as clear as possible.

The cross-sectional independence of the data is maintained for simplicity, one could however allow some forms of dependence in the cross-section. For example, one could allow the type of clustering structure used in Connor et al. (2012). Our results would go through with minimal changes in this case. An alternative approach is to follow Connor and Korajczyk (1993) and to assume that there exists some ordering of the observations with respect to which the data $\{(X_{it}, \varepsilon_{it})\}$ are mixing across $i$. Jenish (2012) derives pointwise limit theorems for nonparametric regression with near-epoch dependent mixing processes defined on a general lattice dimension $d$, which includes that setting as a special case. Robinson (2011) has proposed an alternative approach based on linear processes that does not need a measure of cross-sectional distance. His framework allows for strongly dependent and nonstationary regression disturbances. These types of cross-sectional dependence are much harder to deal with in our framework and would involve a great deal of technical and notational effort to cope with. Heuristically speaking, however, we expect these dependence structures to have no effect on the asymptotic behaviour of our estimators provided the dependence is weak. Specifically, the cross-sectional dependence should wash out of the distribution for the nonparametric estimates and should not affect the univariate asymptotics for the loading coefficients.

We may also allow for nonstationarity in $\{(X_{it}, \varepsilon_{it})\}$ of the type proposed in Dahlhaus (1997). This so-called local stationarity may arise in the time direction, that is, densities change smoothly in the argument $t/T$. In addition, it may arise in the cross-section, that is, densities change smoothly in the argument $i/n$ with respect to an unknown ordering of the individuals. Vogt (2012) establishes a number of results for nonparametric regression with locally stationary processes, and we anticipate that his results can be extended to this case, although the technical effort to accomplish this would be considerable.

It is worth mentioning that our assumptions do not only allow for time series dependence but also for heteroskedasticity in the error terms $\varepsilon_{it}$. The errors may for example have the form $\varepsilon_{it} = \sigma(X_{it})\eta_{it}$, where $\eta_{it}$ are i.i.d. variables independent of $X_{it}$ and $\sigma$ is an unknown volatility function. The moment bounds (1.13)–(1.15) on the error terms are needed to derive a couple of uniform convergence results later on. They are modifications of standard assumptions required to derive uniform convergence rates for kernel estimators; cp. for example Assumption 2 in Hansen (2008). They are for instance satisfied when the error terms take the form $\varepsilon_{it} = \sigma(X_{it})\eta_{it}$, where $\eta_{it}$ are i.i.d. with $\mathbb{E}|\eta_{it}|^\theta < \infty$ and $\sigma$ is a continuous function.

Finally, note that there is a trade-off between the moment condition (1.13) in (A4) and the conditions on the relative sample sizes in (A5). For example, if we restrict attention to the case $n = O(T)$, we can do with $\theta > 4$ in condition (A4). The restrictions in (A5) reflect two constraints on the relative sample sizes: Firstly, $T$ needs to be large enough relative to $n$ such that the preliminary estimators are sufficiently precisely estimated. Secondly, $n$ needs to be large enough such that the error terms stemming from the fixed effect transformation can be ignored.

### 1.5.2 Asymptotics for the estimator $\widehat{\mu}$

Our first result characterizes the asymptotic behaviour of the estimator $\widehat{\mu}$. In particular, it shows that $\widehat{\mu}$ uniformly converges to $\mu$ and is asymptotically normal. To formulate it, we define $V(x)$ to be a $K \times K$ matrix with the entries

$$V_{k,l}(x) = \|K\|_2^2 \lim_{n\to\infty} \Big( n \sum_{i=1}^{n} \omega_{ki}\omega_{li} \frac{\sigma_i^2(x)}{f_i(x)} \Big),$$

where $\sigma_i^2(x) = \mathbb{E}[\varepsilon_{it}^2 | X_{it} = x]$.

**Theorem 1.5.1.** *Let (A1)–(A7) together with ($I_W$1)–($I_W$3) be satisfied. Then*

$$\sup_{x \in I_h} \|\widehat{\mu}(x) - \mu(x)\| = O_p\Big(\sqrt{\frac{\log nT}{nTh}}\Big). \tag{1.16}$$

*Here, $I_h = [C_1 h, 1 - C_1 h]$ if our procedure is based on the Nadaraya-Watson smoothers $\widehat{m}_i^{NW}$ and $I_h = [0,1]$ if it is based on the local linear smoothers $\widehat{m}_i^{LL}$. Moreover, for any fixed point $x \in (0,1)$,*

$$\sqrt{nTh}(\widehat{\mu}(x) - \mu(x)) \xrightarrow{d} N(0, \nu(x)) \tag{1.17}$$

*with $\nu(x) = (S^*)^\intercal V(x) S^*$ and $S^*$ being the limit of $S$.*

The first part of the theorem shows that $\widehat{\mu}$ converges to $\mu$ at a fast rate based on the pooled number of observations $nT$. If we set up our estimation procedure with the local linear smoothers $\widehat{m}_i^{\mathrm{LL}}$, the rate is uniform over the whole support $[0,1]$. For the Nadaraya-Watson based procedure in contrast, the rate is only uniform on the subinterval $[C_1 h, 1 - C_1 h]$ which converges to the support $[0,1]$ as the sample size increases. This is due to the fact that the Nadaraya-Watson estimators $\widehat{m}_i^{\mathrm{NW}}$ suffer from slow convergence rates at the boundary of the support.

The second part of the theorem specifies the asymptotic distribution of $\widehat{\mu}$. The asymptotic covariance matrix $\nu(x)$ can be seen to depend on the weights $\omega_{ki}$. The reason for this is as follows: The normalization of the functions $\mu$ depends on the choice of the weighting matrix $W$. In particular, different choices of $W$ generally result in different eigenvalues $\lambda_k = \int \mu_k^2(x) w(x) dx$, i.e., in different values of the $L_2$-norm of the functions $\mu_k$. This becomes reflected in the covariance matrix $\nu(x)$ through its dependence on the weights $\omega_{ki}$. Moreover, note that $\nu(x)$ need not be diagonal in general: If the weighting matrix $W$ is diagonal, then $V(x)$ is a diagonal matrix as well. However, even then the matrix $S^*$ may have a more complicated non-diagonal structure. Hence, the components of $\widehat{\mu}$ are asymptotically mutually correlated in general.

Regarding inference, we propose a simple plug-in method. Let $\widehat{\varepsilon}_{it} = Y_{it}^{\mathrm{fe}} - \widehat{m}_i(X_{it})$ and

$$\widehat{V}_{k,l}(x) = \|K\|_2^2 \; n \sum_{i=1}^n \omega_{ki} \omega_{li} \frac{\widehat{\sigma}_i^2(x)}{\widehat{f}_i(x)},$$

where $\widehat{\sigma}_i^2(x)$ is a local constant or local linear time series regression smoother of $\widehat{\varepsilon}_{it}^2$ on $X_{it}$ and $\widehat{f}_i(x) = T^{-1} \sum_{t=1}^T K_h(X_{it} - x)$ is the time series kernel density estimator of $f_i(x)$. Then, $\widehat{\nu}(x) = \widehat{S}^\intercal \widehat{V}(x) \widehat{S}$ consistently estimates $\nu(x)$, and pointwise confidence intervals based on this are consistent under our assumptions, see Haerdle (1991).

To derive the results of Theorem 1.5.1, we work with the undersmoothing assumption (A6) on the bandwidth $h$. Moreover, we use the same bandwidth both to estimate the average functions $g$ and the matrix $\Sigma$. It is however also possible to employ different bandwidths. In particular, one may use a slightly undersmoothed bandwidth $h_\Sigma$ of the order $(nT)^{-(1/5+\delta)}$ to construct the estimate $\widehat{\Sigma}$ and a bandwidth $h_g$ of the optimal order $(nT)^{-1/5}$ to set up the estimator $\widehat{g}$. Inspecting the proof of Theorem 1.5.1, it is easily seen that in this case

$$\sqrt{nTh_g}(\widehat{\mu}(x) - \mu(x)) = S^\intercal \left[\sqrt{nTh_g}\big(\widehat{g}(x) - g(x)\big)\right] + o_p(1)$$

with

$$\sqrt{nTh_g}\big(\widehat{g}(x) - g(x)\big) \xrightarrow{d} N(B(x), V(x)),$$

where the variance $V(x)$ has already been defined above and the bias term $B(x)$ is given by $B^{\mathrm{NW}}(x)$ and $B^{\mathrm{LL}}(x)$ in the Nadaraya-Watson and the local linear based case, respectively. The latter two expressions are defined by

$$\begin{aligned}
B_k^{\mathrm{NW}}(x) &= \frac{c_0\mu_2(K)}{2} \lim_{n\to\infty} \sum_{i=1}^n \omega_{ki}\big(2m_i'(x)f_i'(x) + m_i''(x)f_i(x)\big)/f_i(x) \\
B_k^{\mathrm{LL}}(x) &= \frac{c_0\mu_2(K)}{2} \lim_{n\to\infty} \sum_{i=1}^n \omega_{ki} m_i''(x)
\end{aligned}$$

for $k = 1, \dots, K$, where $c_0$ is the limit of the sequence values $\sqrt{nTh_g^5}$.

Given the above remarks, we suggest a straightforward rule of thumb for bandwidth selection. In particular, we first select the bandwidth $h_g$ and then choose the bandwidth $h_\Sigma$ simply by picking a value slightly smaller than the choice of $h_g$. To select the bandwidth $h_g$ (or rather $h_{g,k}$ if we allow a different bandwidth for each function $g_k$), we optimize the integrated mean-squared error criterion

$$\mathrm{IMSE}(h_{g,k}) = h_{g,k}^4 \int B_k^2(x)dx + \frac{1}{nTh_{g,k}} \int V_{k,k}(x)dx$$

for $k = 1, \dots, K$. Minimizing with respect to $h_{g,k}$, the optimal bandwidth turns out to be given by

$$h_{g,k}^* = \left(\frac{\int V_{k,k}(x)dx}{4\int B_k^2(x)dx}\right)^{\frac{1}{5}} (nT)^{-1/5}.$$

This expression still depends on some unknown quantities which have to be replaced by estimators. To do so, we apply a simple plug-in rule similar to the methods discussed in Fan and Gijbels (1996).

### 1.5.3 Asymptotics for the parameter estimators $\widehat{\boldsymbol{\beta}}_i$

The next theorem describes the asymptotic properties of the parameter estimates $\widehat{\beta}_i$ for a fixed individual $i$. To state the asymptotic distribution of $\widehat{\beta}_i$, we introduce the shorthands

$$\Gamma_i = \mathbb{E}[\pi(X_{i0})\mu(X_{i0})\mu(X_{i0})^\intercal] \quad \text{and} \quad \Psi_i = \sum_{l=-\infty}^{\infty} \mathrm{Cov}(\chi_{i0}, \chi_{il}),$$

where $\chi_{it} = \{\pi(X_{it})\mu(X_{it}) - \mathbb{E}[\pi(X_{it})\mu(X_{it})]\}\varepsilon_{it} - \mathbb{E}[\pi(X_{it})\mu(X_{it})]m_i(X_{it})$ and $\pi$ is a bounded weighting function.

**Theorem 1.5.2.** *Suppose that all the assumptions of Theorem 1.5.1 are fulfilled and let $\Gamma_i$ have full rank. Then for any fixed $i$,*

$$\sqrt{T}(\widehat{\beta}_i - \beta_i) \xrightarrow{d} N\big(0, \Gamma_i^{-1}\Psi_i(\Gamma_i^{-1})^\intercal\big).$$

If our procedure is based on Nadaraya-Watson smoothers, we have to restrict the weighting function $\pi$ to equal zero within the boundary region $[0, C_1h) \cup (1-C_1h, 1]$. This is necessary because the convergence rate of $\widehat{\mu}$ is only uniform over the interval $[C_1h, 1-C_1h]$ in this case. If the local linear based procedure is applied, we do not have to impose any restrictions on $\pi$.

From the proof of Theorem 1.5.2, we can see that our parameter estimators $\widehat{\beta}_i$ have some type of oracle property. In particular, it holds that $\sqrt{T}(\widehat{\beta}_i - \widetilde{\beta}_i) = o_p(1)$. Our estimators $\widehat{\beta}_i$ thus have the same asymptotic distribution as the oracle estimators $\widetilde{\beta}_i$ which are constructed under the assumption that the functions $\mu_1, \dots, \mu_K$ are known. To estimate the asymptotic variance $\Psi_i$, we may apply standard long-run variance estimation procedures to the residuals $\widehat{\chi}_{it}$ given by

$$\widehat{\chi}_{it} = \{\pi(X_{it})\widehat{\mu}(X_{it}) - \widehat{\pi\mu}\}\widehat{\varepsilon}_{it} - \widehat{\pi\mu}\widehat{m}_i^e(X_{it}),$$

where we define $\widehat{\pi\mu} = T^{-1}\sum_{t=1}^T \pi(X_{it})\widehat{\mu}(X_{it})$, $\widehat{\varepsilon}_{it} = Y_{it}^{\mathrm{fe}} - \widehat{m}_i^e(X_{it})$ and $\widehat{m}_i^e(x) = \widehat{\beta}_i^\intercal \widehat{\mu}(x)$.

### 1.5.4 Asymptotics for the estimators $\widehat{m}_i^e$ and a parameter of interest

We finally discuss the asymptotic properties of the estimator $\widehat{m}_i^e(x) = \widehat{\beta}_i^\intercal\widehat{\mu}(x)$. It holds that

$$\widehat{m}_i^e(x) - m_i(x) = (\widehat{\beta}_i - \beta_i)^\intercal \mu(x) + \beta_i^\intercal(\widehat{\mu}(x) - \mu(x)) + o_p\Big(\frac{1}{\sqrt{nT}}\Big). \tag{1.18}$$

The first term on the right-hand side is of the order $T^{-1/2}$, while the second one has the (pointwise) order $(nTh)^{-1/2}$ under our conditions. Given assumption (A5) on the relationship between the dimensions $n$ and $T$, the leading term is the first one of order $T^{-1/2}$. It follows that $\widehat{m}_i^e(x)$ is asymptotically normal at the rate $T^{-1/2}$, i.e., at a faster rate than the preliminary estimator $\widehat{m}_i(x)$ which converges at the (pointwise) rate $(Th)^{-1/2}$.

In some empirical applications, a parameter of interest is $c_i = m_i(1) - m_i(0)$. Defining $\widehat{c}_i = \widehat{m}_i^e(1) - \widehat{m}_i^e(0)$, we obtain that

$$\widehat{c}_i - c_i = (\widehat{\beta}_i - \beta_i)^\intercal(\mu(1) - \mu(0)) + \beta_i^\intercal(\widehat{\mu}(1) - \mu(1)) - \beta_i^\intercal(\widehat{\mu}(0) - \mu(0)) + o_p\Big(\frac{1}{\sqrt{nT}}\Big).$$

Under the null hypothesis that $c_i = 0$, we should observe that

$$\sqrt{T}\widehat{c}_i \xrightarrow{d} N(0, \tau_i) \quad \text{with} \quad \tau_i = (\mu(1) - \mu(0))^\intercal \Gamma_i^{-1}\Psi_i(\Gamma_i^{-1})^\intercal(\mu(1) - \mu(0)),$$

which could form the basis of a test. Specifically, we can use the strategy to estimate the covariance matrix $\Gamma_i^{-1}\Psi_i(\Gamma_i^{-1})^\intercal$ from the previous subsection together with the estimators $\widehat{\mu}$ to obtain a consistent estimator $\widehat{\tau}_i$ of the asymptotic variance $\tau_i$ and let

$$t_i = \frac{\widehat{c}_i}{\sqrt{\widehat{\tau}_i/T}},$$

which is asymptotically standard normal.

## 1.6 Robustness of the estimation method

So far, we have worked under the simplifying assumption that the number $K$ of common component functions $\mu_1, \dots, \mu_K$ is known. We now drop this assumption and take into account that $K$ is usually not observed in applications. We only suppose that there is some known upper bound $\overline{K}$ of the number of component functions. In what follows, we investigate how our procedure behaves if we work with this upper bound instead of the true number of components.

To do so, let $\overline{W} = (\overline{\omega}_{ki})$ be a $\overline{K} \times n$ weighting matrix satisfying ($\mathrm{I_W}$3). Writing $\overline{g} = \overline{W}m$ and $\overline{S} = \overline{W}B$, we obtain that

$$\overline{g} = \overline{S}\,\mu.$$

Using an analogous normalization as in Section 1.3, we can assume that (i) the matrix $\int \mu(x)\mu(x)^\intercal w(x)dx$ is diagonal with positive and distinct diagonal entries and that (ii) $\overline{S}$ is a $\overline{K} \times K$ matrix with orthonormal columns. Note that this normalization is somewhat different from that used in the previous sections as we have replaced the weighting scheme $W$ by $\overline{W}$. For simplicity, we suppress this difference in the notation in what follows and again denote the normalized component functions by $\mu$. We thus obtain that

$$\mu = \overline{S}^\intercal \overline{g}.$$

As in the case with known $K$, the matrix $\overline{S}$ can be characterized by an eigenvalue decomposition of the $\overline{K} \times \overline{K}$ matrix

$$\overline{\Sigma} = \int \overline{g}(x)\overline{g}(x)^\intercal w(x)dx.$$

In particular, it holds that $\overline{\Sigma} = \overline{S}D\overline{S}^\intercal$ with $D = \int \mu(x)\mu(x)^\intercal w(x)dx$. Note that this way of writing the spectral decomposition implicitly presupposes that $K$ is known. For this reason, it is more appropriate to rewrite the decomposition as $\overline{\Sigma} = \overline{U}\,\overline{D}\,\overline{U}^\intercal$. Here, $\overline{U}$ is an orthonormal $\overline{K} \times \overline{K}$ matrix with the first $K$ columns being equal to $\overline{S}$. Moreover, $\overline{D} = \int \overline{\mu}(x)\overline{\mu}(x)^\intercal w(x)dx$ is a diagonal $\overline{K} \times \overline{K}$ matrix with $\overline{\mu} = (\mu, 0, \dots, 0)$ being a vector of length $\overline{K}$. Similarly to the case with known $K$, we assume that $\overline{\Sigma}$ converges to a matrix $\overline{\Sigma}^*$ of rank $K$.

To estimate the vector of functions $\overline{\mu} = (\mu, 0, \dots, 0)$, we mimic the estimation procedure from Subsection 1.4.2. In particular, we proceed as follows:

Step 1: Estimate the function $\overline{g}_k(x)$ by $\widetilde{g}_k(x) = \sum_{i=1}^n \overline{\omega}_{ki}\widehat{m}_i(x)$ for $k = 1 \dots, \overline{K}$.

Step 2: Estimate the matrix $\overline{\Sigma}$ by $\widetilde{\Sigma} = \int \widetilde{g}(x)\widetilde{g}(x)^\intercal w(x)dx$.

Step 3: Perform an eigenvalue decomposition of $\widetilde{\Sigma}$ to obtain estimators of $\overline{U}$ and $\overline{D}$. In particular, write $\widetilde{\Sigma} = \widetilde{U}\widetilde{D}\widetilde{U}^\intercal$ with $\widetilde{D}$ being diagonal and $\widetilde{U}$ being orthonormal.

Step 4: Estimate the vector of functions $\overline{\mu} = (\mu, 0, \dots, 0)$ by

$$\widetilde{\mu} = \widetilde{U}^\intercal \widetilde{g}.$$

Inspecting the proof of Theorem 1.5.1, it is straightforward to see that for $k = 1, \dots, K$, the estimator $\widetilde{\mu}_k$ has analogous asymptotic properties as $\widehat{\mu}_k$. In particular, it uniformly converges to $\mu_k$ and is asymptotically normal. The next theorem summarizes the properties of $\widetilde{\mu}_k$ for $k = 1, \dots, K$. To formulate it, we let $\overline{V}(x)$ be a $\overline{K} \times \overline{K}$ matrix with the entries

$$\overline{V}_{k,l}(x) = \|K\|_2^2 \lim_{n\to\infty} \Big( n \sum_{i=1}^n \overline{\omega}_{ki}\overline{\omega}_{li} \frac{\sigma_i^2(x)}{f_i(x)} \Big),$$

where $\overline{\omega}_{ki}$ are the elements of the weighting matrix $\overline{W}$.

**Theorem 1.6.1.** *Let (A1)–(A7) be fulfilled. Then it holds that*

$$\sup_{x \in I_h} \big|\widetilde{\mu}_k(x) - \mu_k(x)\big| = O_p\Big(\sqrt{\frac{\log nT}{nTh}}\Big) \tag{1.19}$$

*for all $k = 1, \dots, K$. As before, $I_h = [C_1 h, 1 - C_1 h]$ for the Nadaraya-Watson based case and $I_h = [0, 1]$ for the local linear based procedure. Moreover, for any fixed point*

$x \in (0,1)$,

$$\sqrt{nTh}\,[\widetilde{\mu}(x) - \mu(x)] \xrightarrow{d} N(0, \overline{\nu}(x)), \tag{1.20}$$

*where* $\overline{\nu}(x) = (\overline{S}^*)^\intercal \overline{V}(x)\, \overline{S}^*$ *and* $\overline{S}^*$ *is the limit of* $\overline{S}$.

In addition, we can show that for $k = K+1, \ldots, \overline{K}$, the estimators $\widetilde{\mu}_k$ converge in an $L_2$-sense to zero.

**Theorem 1.6.2.** *Let (A1)–(A7) be fulfilled. Then it holds that*

$$\int \widetilde{\mu}_k^2(x) w(x) dx = o_p\Big(\frac{1}{\sqrt{nTh}}\Big) \tag{1.21}$$

*for all* $k = K+1, \ldots, \overline{K}$.

The proof of Theorem 1.6.2 is given in Appendix A.1. Taken together, Theorems 1.6.1 and 1.6.2 show that our procedure is robust to overestimating the number of component functions $K$. In particular, applying it with the upper bound $\overline{K}$ instead of $K$, the first $K$ components of the estimator $\widetilde{\mu}$ still uniformly converge to the vector of functions $\mu$. Moreover, the remaining components converge to zero in an $L_2$-sense and thus become negligible as the sample size grows.

## 1.7 Selecting the number of components $K$

In this section, we propose a simple method to estimate the unknown number of components $K$. To define our estimator, let $\overline{\lambda} = (\overline{\lambda}_1, \ldots, \overline{\lambda}_{\overline{K}})^\intercal$ be the vector of eigenvalues of the matrix $\overline{\Sigma}$ arranged in descending order. Analogously, let $\widetilde{\lambda}$ be the eigenvalues of the estimator $\widetilde{\Sigma}$. Finally, let $\{\delta_{n,T}\}$ be any null sequence of positive numbers which converges to zero at the order $O(1/\sqrt{nTh})$ or at a slower rate. With this notation at hand, our estimator of $K$ is defined as

$$\widehat{K} = \min\left\{ k \in \{1, \ldots, \overline{K}\} \;\middle|\; \frac{\widetilde{\lambda}_1 + \ldots + \widetilde{\lambda}_k}{\widetilde{\lambda}_1 + \ldots + \widetilde{\lambda}_{\overline{K}}} \geq 1 - \delta_{n,T} \right\}.$$

The intuition behind this estimator is simple: Under our assumptions, the matrix $\overline{\Sigma}$ has $K$ non-zero eigenvalues, i.e., the first $K$ entries of $\overline{\lambda}$ are non-zero. The first $K$ entries of the estimator $\widetilde{\lambda}$ thus converge to some positive values, whereas the other ones approach zero as the sample size increases. Hence, the ratio

$$\frac{\widetilde{\lambda}_1 + \ldots + \widetilde{\lambda}_k}{\widetilde{\lambda}_1 + \ldots + \widetilde{\lambda}_{\overline{K}}}$$

should converge to a number strictly smaller than 1 for $k < K$ and to 1 for $k \geq K$. This suggests that $\widehat{K}$ consistently estimates the true number of components $K$.

This intuition can easily be turned into a formal argument: First of all, it can be shown that the convergence rate of $\widetilde{\lambda}$ is at least $o_p(1/\sqrt{nTh})$, i.e., $\|\widetilde{\lambda} - \overline{\lambda}\| = o_p(1/\sqrt{nTh})$. As a consequence, it holds that

$$\frac{\widetilde{\lambda}_1 + \ldots + \widetilde{\lambda}_k}{\widetilde{\lambda}_1 + \ldots + \widetilde{\lambda}_{\overline{K}}} = \frac{\lambda_1 + \ldots + \lambda_k}{\lambda_1 + \ldots + \lambda_{\overline{K}}} + o_p\Big(\frac{1}{\sqrt{nTh}}\Big).$$

for any $k \in \{1, \ldots, \overline{K}\}$. In particular,

$$\frac{\widetilde{\lambda}_1 + \ldots + \widetilde{\lambda}_K}{\widetilde{\lambda}_1 + \ldots + \widetilde{\lambda}_{\overline{K}}} = 1 + o_p\Big(\frac{1}{\sqrt{nTh}}\Big).$$

Using these two equations together with some straightforward arguments, it is easily seen that $\widehat{K}$ is indeed a consistent estimator of the true number of components $K$, i.e. $\widehat{K} = K + o_p(1)$.

When implementing the estimator $\widehat{K}$ in practice, an important question is how to choose the constant $\delta_{n,T}$. We suggest to pick it by a rule of thumb which is similar to the procedure usually used in principal component analysis for selecting the number of factors. To understand the intuitive idea behind the rule, first note that $\lambda_k = \int \mu_k^2(x) w(x) dx$ for $k = 1, \ldots, K$ and $\lambda_k = 0$ for $k = K+1, \ldots, \overline{K}$. The eigenvalues $\lambda_k$ are thus equal to (the square of) a weighted $L_2$-norm of the component functions $\overline{\mu} = (\mu, 0, \ldots, 0)$. Put differently, they measure the variation of these functions. As a result, the ratio

$$\frac{\lambda_1 + \ldots + \lambda_k}{\lambda_1 + \ldots + \lambda_{\overline{K}}}$$

can be interpreted to capture the percentage of the overall variation in the functions $\overline{\mu}$ that stems from the first $k$ components. Hence, by picking a certain value of $\delta_{n,T}$, we select the number of component functions such that at least a certain percentage of the overall variation is explained by the chosen number of components. For instance, if we let $\delta_{n,T} = 0.05$, we pick the number of components to capture at least 95% of the total variation. Keeping in mind that our estimation procedure is robust to picking the number of components too large, we propose to choose the constant $\delta_{n,T}$ rather small (e.g. $\delta_{n,T} = 0.01$ or $\delta_{n,T} = 0.05$). This results in a conservative rule which tends to overestimate the true number $K$ rather than to underestimate it. As already noted above, this way of selecting the number of components is very similar to the usual approach in factor analysis (see e.g. Zhu and Ghodsi (2006) or chapter 6 of Jolliffe (2002)).

## 1.8 Simulation study

To assess the small sample properties of our estimation methods, we simulate data from the following model setup: The regressors $X_{it}$ are i.i.d. draws from a uniform distribution on the unit interval. Moreover, there are $K = 2$ common component functions defined by

$$\mu_1(x) = \sqrt{2}\sin(2\pi x) \qquad \text{and} \qquad \mu_2(x) = \sqrt{2}\cos(2\pi x).$$

These functions are orthonormal with respect to the standard scalar product on $[0, 1]$, i.e., $\int_0^1 \mu_1(x)\mu_2(x)dx = 0$ and $\int_0^1 \mu_k^2(x)dx = 1$ for $k = 1, 2$. As the regressors are uniformly distributed on $[0, 1]$, we obtain that $\mathbb{E}[\mu_k(X_{it})] = 0$ for $k = 1, 2$ and thus $\mathbb{E}[m_i(X_{it})] = 0$ with $m_i(x) = \beta_{i1}\mu_1(x)+\beta_{i2}\mu_2(x)$. Thus, the regression functions fulfill the normalization $\mathbb{E}[m_i(X_{it})] = 0$ that is assumed for identification.

The factor loadings $\beta_{ik}$ ($i = 1, \ldots, n$, $k = 1, \ldots, K$) are generated deterministically according to

$$\beta_{i1} = 1 + \frac{i-1}{n-1} \qquad \text{and} \qquad \beta_{i2} = 2 - \frac{i-1}{n-1}.$$

With this choice, the coefficient $\beta_{i1}$ of the function $\mu_1$ linearly increases from 1 to 2 as the index $i$ grows larger. Similarly, the loading $\beta_{i2}$ of $\mu_2$ decreases from 2 to 1. Hence, the component function $\mu_1$ becomes more and more important as the index $i$ gets larger and vice versa for the second component $\mu_2$. The weighting matrix $W$ is given by

$$W = \begin{pmatrix} 2/n & \ldots & 2/n & 0 & \ldots & 0 \\ 0 & \ldots & 0 & 2/n & \ldots & 2/n \end{pmatrix}.$$

Note that the coefficient matrix $B$ and the weighting matrix $W$ are chosen such that $S = WB$ has full rank. In addition, the $\mu$-functions are orthonormal. Hence, the normalization conditions of Section 1.3 are fulfilled. In the simulations, $S$ and $\mu$ are re-normalized such that they fulfill condition ($\mathrm{I_W}1$).

The individual and time fixed effects $\alpha_i$ and $\gamma_t$ are i.i.d. standard normal random variables. The model constant $\mu_0$ is set to zero, and the disturbances $\varepsilon_{it}$ are i.i.d. normal random variables with zero mean and standard deviation $\sigma_\varepsilon$. To vary the signal-to-noise ratio in the model, we choose two different values for $\sigma_\varepsilon$, in particular $\sigma_\varepsilon \in \{1, 2\}$. As can be seen, there is no time series dependence in the error terms and the regressors, and we have only included a very limited form of fixed effects. These simplifications allow us to get a clear picture of the performance of our estimation methods. It goes without saying that they may be relaxed, i.e., we may allow for time series dependence in the model variables and add some more complicated forms of fixed effects.

In what follows, we examine the performance of our estimators $\widehat{\mu}$ and $\widehat{\beta}_i$.

Moreover, we assess the small sample behaviour of two estimators of the average regression function $m_{\text{av}}(x) = n^{-1}\sum_{i=1}^n m_i(x)$ defined by $\widehat{m}_{\text{av}}(x) = n^{-1}\sum_{i=1}^n \widehat{m}_i(x)$ and $\widehat{m}^e_{\text{av}}(x) = n^{-1}\sum_{i=1}^n \widehat{m}^e_i(x)$, where $\widehat{m}^e_i(x) = \widehat{\beta}_i^\intercal \widehat{\mu}(x)$ are the reconstructed regression functions. As performance measures, we employ the mean squared errors

$$\text{MSE}(\widehat{\mu}_k) = \int_0^1 \big[\widehat{\mu}_k(x) - \mu_k(x)\big]^2 dx$$

for $k = 1, 2$ along with

$$\begin{aligned}\text{MSE}(\widehat{m}_{\text{av}}) &= \int_0^1 \big[\widehat{m}_{\text{av}}(x) - m_{\text{av}}(x)\big]^2 dx \\ \text{MSE}(\widehat{m}^e_{\text{av}}) &= \int_0^1 \big[\widehat{m}^e_{\text{av}}(x) - m_{\text{av}}(x)\big]^2 dx.\end{aligned}$$

The small sample behavior of the coefficient estimates $\widehat{\beta}_i$ is evaluated by the $L_1$-norm

$$\frac{1}{n}\sum_{i=1}^n |\widehat{\beta}_{ik} - \beta_{ik}|$$

for $k = 1, 2$. Throughout, we assume the number of components $K = 2$ to be known and use the version of our method which is based on local linear estimators. Moreover, the bandwidth is set to $h = 0.15$ and we use an Epanechnikov kernel. As a robustness check, we have varied the bandwidth. As this produces very similar results, we have however not reported them here. Finally, the number of replications is set to $N = 1000$.

Tables 1.1 and 1.2 report the simulation results. Overall, our estimators perform well even for the moderate sample sizes $n = T = 50$. The accuracy of the estimators increases steadily as the dimensions $n$ and $T$ grow larger, the only exception being the estimates of the factor loadings which improve above all in $T$ but not so much in $n$. This is a very natural phenomenon as the factor loadings are estimated from individual time series regressions. Hence, their quality should depend above all on the time series dimension and not so much on the length of the cross-section. It is also worth mentioning that the MSE of the reconstructed average $\widehat{m}^e_{\text{av}}$ is smaller and converges faster to zero than the MSE of $\widehat{m}_{\text{av}}$. This observation is consistent with the asymptotic properties of the estimators $\widehat{m}_i$ and $\widehat{m}^e_i$: While $\widehat{m}_i$ converges at the rate $(Th)^{-1/2}$, $\widehat{m}^e_i$ converges at the faster rate $T^{-1/2}$ (cp. Section 1.5.4). Finally, when the standard deviation $\sigma_\varepsilon$ of the disturbance terms is increased to 2, the signal-to-noise ratio in the model decreases. This makes it harder to estimate the functions and parameters of interest, which is reflected in higher values of the MSE and the $L_1$-norm as can be seen upon comparing Tables 1.1 and 1.2.

## 1.9 Conclusion

Our model captures in a general way two important features in many applications: heterogeneity and nonlinearity. We also allow for a limited type of endogeneity through the unobserved time and cross-section fixed effects. Nevertheless, our estimation procedures are particularly simple, and are in fact closed form at each step. We have provided the tools to conduct inference and to select tuning and order parameters.

We close the chapter by commenting on some extensions of our model. In our analysis, we have focused on the case of univariate regressors $X_{it}$. If the regressors are multivariate, the usual curse of dimensionality problem arises, cp.Stone (1980). One way to circumvent this problem is to assume that the regression functions $m_i$ split up into additive components according to

$$m_i(x) = m_i^{(1)}(x_1) + \ldots + m_i^{(d)}(x_d),$$

where $d$ is the dimension of the regressors. Analogously to the univariate case, we may suppose that for each $j$, the individual functions $m_i^{(j)}$ have the common component structure

$$m_i^{(j)}(x_j) = \sum_{k=1}^{K} \beta_{ik}^{(j)} \mu_k^{(j)}(x_j),$$

where $K$ could also be allowed to differ across $j$. The additive functions $m_i^{(1)}, \ldots, m_i^{(d)}$ can be estimated by time series backfitting for each individual $i$, see Mammen et al. (1999). These backfitting estimators may be used as preliminary estimators in our procedure. In particular, the common functions $\mu^{(j)} = (\mu_1^{(j)}, \ldots, \mu_K^{(j)})$ may be estimated separately for each $j$ by repeating the estimation steps of Section 1.4 based on the backfitting estimators.

Perhaps one is also concerned that we do not allow for sufficiently general time effects, since we have assumed homogeneous such effects. A more general model which allows for additional interactive (exogenous) time effects is given by

$$Y_{it} = \mu_0 + \alpha_i + \gamma_t + g_i(t/T) + m_i(X_{it}) + \epsilon_{it},$$

where $g_i(\cdot)$ is a smooth function of rescaled time. In practice, a number of authors adopt parametric specifications for $g_i(t/T)$ such as $g_i(t/T) = \zeta_i t + \eta_i t^2$, see for example Brogaard et al. (2013). In this case, we obtain

$$Y_{it}^{\text{fe}} = g_i(t/T) + m_i(X_{it}) + \epsilon_{it} + O_p(T^{-1/2}) + O_p(n^{-1/2}),$$

where we have assumed that $\sum_{t=1}^{T} g_i(t/T) = 0$. Similarly to the multivariate case discussed above, we here have an additive regression model that could be estimated by time series backfitting. Moreover, one could restrict $g_i(\cdot)$ to rely on a small number of principal components as we do for $m_i(\cdot)$, and do parallel analysis for both functions.

**Table 1.1:** Small sample properties of the estimators in the design with $\sigma_\varepsilon = 1$

*a) MSE of* $\widehat{m}_{av}$

| $T\backslash n$ | 50 | 100 | 150 | 200 |
|---|---|---|---|---|
| 50 | 0.0449 | 0.0362 | 0.0331 | 0.032 |
| 100 | 0.0425 | 0.0347 | 0.0324 | 0.0311 |
| 150 | 0.042 | 0.0345 | 0.0321 | 0.0309 |
| 200 | 0.0418 | 0.0343 | 0.0321 | 0.0308 |

*b) MSE of* $\widehat{m}^e_{av}$

| $T\backslash n$ | 50 | 100 | 150 | 200 |
|---|---|---|---|---|
| 50 | 0.017 | 0.0124 | 0.0107 | 0.0101 |
| 100 | 0.0109 | 0.0077 | 0.0071 | 0.0066 |
| 150 | 0.0092 | 0.0067 | 0.0061 | 0.0058 |
| 200 | 0.0085 | 0.0062 | 0.0057 | 0.0054 |

*c) MSE of* $\widehat{\mu}_1$

| $T\backslash n$ | 50 | 100 | 150 | 200 |
|---|---|---|---|---|
| 50 | 0.0159 | 0.0099 | 0.0064 | 0.0052 |
| 100 | 0.008 | 0.004 | 0.003 | 0.0024 |
| 150 | 0.0053 | 0.0029 | 0.0021 | 0.0016 |
| 200 | 0.0041 | 0.0022 | 0.0016 | 0.0012 |

*d) MSE of* $\widehat{\mu}_2$

| $T\backslash n$ | 50 | 100 | 150 | 200 |
|---|---|---|---|---|
| 50 | 0.0159 | 0.0092 | 0.0063 | 0.005 |
| 100 | 0.009 | 0.0051 | 0.0039 | 0.0035 |
| 150 | 0.0065 | 0.0043 | 0.0035 | 0.003 |
| 200 | 0.0054 | 0.0035 | 0.003 | 0.0027 |

*e)* $L_1$*-norm of the coefficient estimates* $\widehat{\beta}_{i1}$

| $T\backslash n$ | 50 | 100 | 150 | 200 |
|---|---|---|---|---|
| 50 | 0.129 | 0.125 | 0.124 | 0.123 |
| 100 | 0.089 | 0.0853 | 0.0841 | 0.0837 |
| 150 | 0.072 | 0.0684 | 0.0679 | 0.0675 |
| 200 | 0.0627 | 0.0591 | 0.0583 | 0.0581 |

*f)* $L_1$*-norm of the coefficient estimates* $\widehat{\beta}_{i2}$

| $T\backslash n$ | 50 | 100 | 150 | 200 |
|---|---|---|---|---|
| 50 | 0.136 | 0.13 | 0.128 | 0.128 |
| 100 | 0.0973 | 0.0914 | 0.0895 | 0.0886 |
| 150 | 0.0822 | 0.0752 | 0.0732 | 0.0721 |
| 200 | 0.0732 | 0.0658 | 0.0641 | 0.0629 |

**Table 1.2:** Small sample properties of the estimators in the design with $\sigma_\varepsilon = 2$

*a) MSE of $\widehat{m}_{av}$*

| $T \backslash n$ | 50 | 100 | 150 | 200 |
|---|---|---|---|---|
| 50 | 0.0512 | 0.039 | 0.0352 | 0.034 |
| 100 | 0.0456 | 0.0362 | 0.0334 | 0.0318 |
| 150 | 0.0442 | 0.0354 | 0.0327 | 0.0314 |
| 200 | 0.0428 | 0.035 | 0.0326 | 0.0312 |

*b) MSE of $\widehat{m}^e_{av}$*

| $T \backslash n$ | 50 | 100 | 150 | 200 |
|---|---|---|---|---|
| 50 | 0.0233 | 0.0153 | 0.0129 | 0.0118 |
| 100 | 0.0144 | 0.00936 | 0.008 | 0.00749 |
| 150 | 0.0115 | 0.00779 | 0.00682 | 0.00632 |
| 200 | 0.00993 | 0.0069 | 0.00634 | 0.00579 |

*c) MSE of $\widehat{\mu}_1$*

| $T \backslash n$ | 50 | 100 | 150 | 200 |
|---|---|---|---|---|
| 50 | 0.0343 | 0.019 | 0.0129 | 0.0103 |
| 100 | 0.0169 | 0.0089 | 0.00604 | 0.00465 |
| 150 | 0.0106 | 0.0057 | 0.00402 | 0.00294 |
| 200 | 0.00804 | 0.00418 | 0.00292 | 0.00225 |

*d) MSE of $\widehat{\mu}_2$*

| $T \backslash n$ | 50 | 100 | 150 | 200 |
|---|---|---|---|---|
| 50 | 0.0339 | 0.0183 | 0.0125 | 0.00993 |
| 100 | 0.0171 | 0.00942 | 0.0071 | 0.00568 |
| 150 | 0.0117 | 0.007 | 0.00542 | 0.00429 |
| 200 | 0.00955 | 0.00555 | 0.00433 | 0.00366 |

*e) $L_1$-norm of the coefficient estimates $\widehat{\beta}_{i1}$*

| $T \backslash n$ | 50 | 100 | 150 | 200 |
|---|---|---|---|---|
| 50 | 0.233 | 0.231 | 0.231 | 0.231 |
| 100 | 0.162 | 0.162 | 0.161 | 0.161 |
| 150 | 0.134 | 0.131 | 0.131 | 0.131 |
| 200 | 0.115 | 0.113 | 0.114 | 0.114 |

*f) $L_1$-norm of the coefficient estimates $\widehat{\beta}_{i2}$*

| $T \backslash n$ | 50 | 100 | 150 | 200 |
|---|---|---|---|---|
| 50 | 0.237 | 0.234 | 0.234 | 0.234 |
| 100 | 0.169 | 0.166 | 0.164 | 0.164 |
| 150 | 0.138 | 0.135 | 0.134 | 0.134 |
| 200 | 0.12 | 0.118 | 0.116 | 0.116 |

# Chapter 2

# The Effect of Fragmentation in Trading on Market Quality in the UK Equity Market[1]

We investigate the effects of fragmentation in equity markets on the quality of trading outcomes in a panel of FTSE stocks over the period 2008-2011. This period coincided with a great deal of turbulence in the UK equity markets which had multiple causes that need to be controlled for. To achieve this, we use the common correlated effects estimator for large heterogeneous panels. We extend this estimator to quantile regression to analyze the whole conditional distribution of market quality. We find that both fragmentation in visible order books and dark trading that is offered outside the visible order book lower volatility. But dark trading increases the variability of volatility and trading volumes. Visible fragmentation has the opposite effect on the variability of volatility in particular at the upper quantiles of the conditional distribution. The transition from a monopolistic to a fragmented market is non-monotonic with respect to the degree of fragmentation.

[1]This chapter is written in jointly with Oliver Linton and Michael Vogt and a version of it is forthcoming in the *Journal of Applied Econometrics*.

## 2.1 Introduction

The implementation of the "Markets in Financial Instruments Directive (MiFID)" has had a profound impact on the organization of security exchanges in Europe. Most importantly, it abolished the concentration rule in European countries that required all trading to be conducted on primary exchanges and it created a competitive environment for equity trading; new types of trading venues that are known as Multilateral Trading Facilities (MTF) or Systematic Internalizers (SI) were created that fostered this competition. As a result, MiFID has served as a catalyst for the competition between equity marketplaces we observe today. The first round of MiFID was implemented across Europe on November 1st, 2007, although fragmentation of the UK equity market began sometime before that (since the UK did not have a formal concentration rule), and by 13th July, 2007, Chi-X was actively trading all of the FTSE 100 stocks. In April 2014, the volume of the FTSE 100 stocks traded via the London Stock Exchange (LSE) had declined to 51%.[2] Similar developments have taken place across Europe.

At the same time, there has been a trend towards industry consolidation: a number of mergers of exchanges allowed cost reductions through "synergies" and also aided standardization and pan European trading. For example, Chi-X was acquired by BATS in 2011. There are reasons to think that consolidation fosters market quality. A single, consolidated exchange market creates network externalities. In addition, some have argued that security exchanges even qualify as natural monopolies. On the other hand, there are arguments for why competition between trading venues can improve market quality. Higher competition generally promotes technological innovation, improves efficiency and reduces the fees that have to paid by investors. Furthermore, traders that use Smart Order Routing Technologies (SORT) can still benefit from network externalities in a fragmented market place.

In view of the ambiguous theoretical predictions, whether the net effect of fragmentation on market quality is negative or positive is an empirical question. In this chapter, we investigate the effect of visible fragmentation and dark trading on measures of market quality such as volatility, liquidity, and trading volume in the UK equity market. Our analysis distinguishes between the effect of fragmentation on average market quality on the one hand and on its variability on the other hand. The first question sheds light on the relationship between fragmentation and market quality during "normal" times. In contrast, the second question investigates whether fragmentation of trading has led to an increase in the frequency of liquidity droughts or to more extraordinary price moves. This latter issue has been raised in several studies that have analyzed the Flash Crash and other recent market meltdowns (Madhavan (2012)). Of course, there is no market structure that can entirely eliminate variability in liquidity or trading volume. But regulators

[2]http://www.batstrading.co.uk/market_data/market_share/index/, accessed on April 16, 2014

aim at constructing a robust market structure that contributes to an orderly and resilient functioning of equity markets in times of market turmoil. One reason for this objective is that investors particularly value the ability to trade in times of market stress and a stable market structure is thus important to maintain investor confidence (Securities Exchange Commission (2013)).

We use a novel dataset that allows us to calculate weekly measures for overall fragmentation, visible fragmentation and dark trading that is offered outside the visible order book for each firm of the FTSE 100 and FTSE 250 indices and combine it with data on indicators of market quality. To investigate the effect of fragmentation on market quality, we use a version of Pesaran (2006) common correlated effects (CCE) estimator for heterogeneous panels. That estimator is suitable for our data because it can account for common but unobserved factors that affect both fragmentation and market quality. For example, these factors account for the activity of High Frequency Traders (HFT) whose activity has generated so much scrutiny (Foresight (2012)). The unobserved factors also control for the global financial crisis, changes in trading technology or new types of trading strategies. We extend Pesaran (2006) estimator to quantile regression (the QCCE estimator) to analyze the whole conditional distribution of market quality. This estimator is also robust to extreme observations on the response.

We find that overall fragmentation, visible fragmentation and dark trading lower volatility at the LSE. But dark trading increases the variability of volatility and trading volumes. Fragmentation has the opposite effect on the variability of volatility in particular at the upper quantiles of the conditional distribution. This result is robust across alternative measures of variability in market quality. Trading volume both globally and locally at the LSE is higher if visible order books are less fragmented or if there is more dark trading. Compared to a monopolistic market, visible fragmentation lowers liquidity measured by quoted bid-ask spreads at the LSE. We also investigate the transition between monopoly and competition in terms of the level of fragmentation. We find this transition is non-monotonic for overall and visible fragmentation and takes the form of an inverted U shape. The level of optimal fragmentation varies across individual firms but it is positively related to market capitalization.

The remainder of this chapter is organized as follows. Section 2.2 discusses the related literature. The data and measures for fragmentation and market quality are introduced in Section 2.3. Section 2.4 proposes an econometric framework suitable for answering the questions of interest and Section 2.5 reports the results. Section 2.6 concludes.

## 2.2 Related literature

Recently, regulators in both Europe and the US introduced new provisions to modernize and strengthen their financial markets. The "Regulation of National Markets (RegNMS)" in the US was implemented in 2005, two years earlier than its European counterpart MiFID.[3] One common theme of these regulations is to foster competition between equity trading venues. But RegNMS and MiFID differ in important aspects: under RegNMS, trades and quotes are recorded on an official consolidated tape and trade-throughs are prohibited, while in Europe, a (publicly guaranteed) consolidated tape does not yet exist, and trade-throughs are allowed.[4]

These regulatory changes and institutional differences between Europe and the US have motivated an ongoing debate among academics and practitioners on the effect of competition between trading venues on market quality. The remainder of this section summarizes some theoretical predictions and existing empirical evidence for both Europe and the US.

**Theoretical predictions.** On the one hand, there are theoretical reasons for why competition can harm market quality. Security exchanges may be natural monopolies because a single exchange has lower costs when compared to a fragmented market place. In addition, a single, consolidated exchange market creates network externalities. The larger the market, the more trading opportunities exist that attract even more traders by reducing the execution risk. Theoretical models that incorporate network externalities predict that liquidity should concentrate at one trading venue (Pagano (1989)). This prediction is at odds with the fragmentation of order flow we observe today. One possible explanation is that traders that use SORT can still benefit from network externalities in a fragmented market place. Such a situation is modelled by Foucault and Menkveld (2008) who study the competition between Euronext and the LSE in the Dutch equity market. Before the entry of LSE, the Dutch equity market had a centralized limit order book that was operated by Euronext. Their theory predicts that a larger share of SORT increases the liquidity supply of the entrant.

On the other hand, there are reasons why competition between trading venues can improve market quality. Higher competition generally promotes technological innovation, improves efficiency and reduces the fees that have to be paid by investors.[5] Biais et al. (2000) propose a model of imperfect competition in financial markets that is consistent with the observation that traders earn positive profits and

[3]The different pillars of MiFID are summarized in Appendix B.1.

[4]A trade-through occurs if a sell (buy) order is executed at a price that is higher (lower) than the best price quoted in the market.

[5]For example, the latency at BATS was about 8 to 10 times lower when compared to the LSE in 2010 (Wagener (2011)), and the LSE has responded by upgrading its system at a faster pace (cp. Appendix B.3). Chesini (2012) reports a reduction in explicit trading fees on exchanges around Europe due to the competition between them for order flow.

that the number of traders is finite. Their model also assumes that traders have private information on the value of financial assets, giving rise to an asymmetric information issue. When compared to a monopolistic market, their model predicts that a competitive market is characterized by lower spreads and a higher trading volume. Buti et al. (2010) study the competition between a trading venue with a transparent limit order book and a dark pool. Their model implies that after the entry of the dark pool, the trading volume in the limit order book decreases, while the overall volume increases.

**Empirical evidence for Europe.** After the introduction of MiFID, equity trading in Europe became more fragmented as new trading venues gained significant market shares from primary exchanges. Gresse (2011) investigates if fragmentation of order flow has had a positive or negative effect on market quality in European equity markets. She examines this from two points of view. First, from the perspective of a sophisticated investor who has access to SORT and thus to the consolidated order book. Second, from the point of view of an investor who can only access liquidity on the primary exchange. Her sample covers stocks listed on the LSE and Euronext exchanges in Amsterdam, Paris and Brussels for 1 month in 2007 and 3 months in 2009. Gresse (2011) finds that increased competition between trading venues creates more liquidity both locally and globally, and that dark trading does not have a negative effect on liquidity.

De Jong et al. (2015) study the effect of fragmentation on market quality in a sample of 52 Dutch stocks for the period from 2006 to 2009. They distinguish between platforms with a visible order book and dark platforms that operate an invisible order book. Their primary finding is that fragmentation on trading venues with a visible order book improves global liquidity, but has a negative effect on local liquidity. But visible fragmentation ceases to improve global liquidity when it exceeds a turning point. Dark trading is found to have a negative effect on liquidity.

Studying UK data, Linton (2012) does not find a detrimental effect of fragmentation on volatility using daily data for the FTSE 100 and FTSE 250 indices for the period from 2008 to 2011. Hengelbrock and Theissen (2009) study the market entry of Turquoise in September 2008 in 14 European countries. Their findings suggest that quoted bid-ask spreads on regulated markets declined after the entry of Turquoise. Riordan et al. (2011) also analyze the contribution of the LSE, Chi-X, Turquoise and BATS to price discovery in the UK equity market. They find that the most liquid trading venues LSE and Chi-X dominate price discovery. Over time, the importance of Chi-X in price discovery has increased.

Overall, the evidence for Europe suggests that the positive effects of fragmentation on market quality outweighs its negative effects. A possible reason for the observed improvement in market quality despite the lack of trade-through protection and a consolidated tape are algorithmic traders and HFT (Riordan et al. (2011)).

By relying on SORT, these traders create a virtually integrated marketplace in the absence of a commonly owned central limit order book.

**Empirical evidence for the US.** In contrast to Europe, competition between trading venues is not a new phenomenon in the US where Electronic Communication Networks (ECN) started to compete for order flow already in the 1990s. Boehmer and Boehmer (2003) investigate if the entry of the NYSE into the trading of Exchange Traded Funds (ETFs) has harmed market quality. Prior to the entry of the NYSE, the American Stock Exchange, the Nasdaq InterMarket, regional exchanges and ECNs already traded ETFs. Boehmer and Boehmer (2003) document that increased competition reduced quoted, realized and effective spreads and increased depth.

O'Hara and Ye (2011) analyze the effect of the proliferation of trading venues on market quality for a sample of stocks that are listed on NYSE and Nasdaq between January and June 2008. They find that stocks with more fragmented trading had lower spreads and faster execution times. In addition, fragmentation increases short-term volatility but is associated with greater market efficiency. Drawing on their findings for the US, O'Hara and Ye (2011) hypothesize that trade-through protection and a consolidated tape are important for the emergence of a single virtual market in Europe. This hypothesis is supported by the findings of Foucault and Menkveld (2008). However, Riordan et al. (2011) conclude that the existence of trade-throughs does not harm market quality.

To summarize, the evidence for the US points to an improvement in average market quality in a fragmented market place. Notwithstanding these results on average quality, Madhavan (2012) finds that both trade fragmentation and quote fragmentation prior to the Flash Crash are associated with larger drawdowns during the Flash Crash. This finding suggests that fragmentation may be affecting the variability of market quality. Below, we further investigate this question.

## 2.3 Data and measurement issues

This section discusses how we measure fragmentation, dark trading and market quality. Our data on market quality and fragmentation covers the period from May 2008 to June 2011 and includes all individual FTSE 100 and 250 firms. At the time of writing, the FTSE350 index companies are valued at $3400 billion, which represents a substantial part of the UK (and European) equity market.

### 2.3.1 Fragmentation and dark trading

Weekly data on the volume of the individual firms traded on each equity venue was supplied to us by Fidessa.[6] For venue $j = 1, \ldots, J$, denote by $w_j$ the market share (according to the number of shares traded) of that venue. We measure fragmentation by the dispersal of volume across multiple trading venues, or $1 - \sum w_j^2$, where $\sum w_j^2$ is the Herfindahl index.

In May 2008, equity trading in the UK was consolidated at the LSE as reflected by an average fragmentation level of 0.4 (Figure 2.1). By June 2011, the entry of new trading venues has changed the structure of the UK equity market fundamentally: fragmentation has increased by about half over the sample period. The rise of HFT is one explanation of the successful entry of alternative trading venues. These venues could attract a significant share of HFT order flow by offering competitive trading fees and sophisticated technologies. In particular, MTF's typically adopt the so-called maker-taker rebates that reward the provision of liquidity to the system, allow various new types of orders, and have small tick sizes. Additionally, their computer systems offer a lower latency when compared to regulated markets. This is probably not surprising since MTFs are often owned by a consortium of users, while the LSE is a publicly owned corporation.

The data allow us to distinguish between public exchanges with a visible order book ("lit"), regulated venues with an invisible order book ("regulated dark pools"), over the counter ("OTC") venues, and systematic internalizers ("SI").[7] We use this information in our analysis to distinguish between fragmentation in visible order books (Figure 2.1) and dark trading (Figure 2.2). Following Gresse (2011) and De Jong et al. (2015), dark trading is measured as the share of volume traded on OTC venues, regulated dark pools and SI. The share of these different categories of dark trading increased over the sample period, while the share of volume traded at lit venues has fallen considerably. For all categories, the observed changes are largest in 2009. In the period after 2009, volumes have approximately stabilized with the exception of regulated dark venues where volume kept increasing. Quantitatively, the majority of trades are executed on lit and OTC venues while regulated dark and SI venues attract only about 1% of the order flow.

### 2.3.2 Market quality

We measure market quality by volatility, liquidity, and trading volume of the FTSE 100 and 250 stocks. Since our measure of fragmentation is only available at a weekly frequency, all measures of market quality are constructed as weekly medians of the

[6] In the Appendix B.2, we give a full list of the trading venues in our sample.

[7] Not all trading venues with an invisible order book are registered as dark pools: unregulated categories of dark pools are registered as OTC venues or brokers (Gresse (2012)).

daily measures.[8]

With the exception of trading volume, our measures of market quality are calculated using data from the LSE. In that sense, our measures are local as compared to global measures that are constructed by consolidating measures from all markets. Global measures are relevant for investors that have access to SORT, while local measures are important for small investment firms that are only connected to the primary exchange (perhaps to save costs) or for retail investors that are restricted by the best execution policy of their investment firm.[9] For example, Gomber et al. (2012) provide evidence that 20 out of 75 execution policies in their sample state that they only execute orders at the primary exchange.

**Volatility.** Volatility is often described in negative terms, but its interpretation should depend on the perspective and on the type of volatility.[10] For example, Bartram et al. (2012) argue that volatility levels in the US are in many respects higher than in other countries but this reflects more innovation and competition rather than poor market quality.

One well known method to estimate volatility is due to Parkinson (1980). The Parkinson estimator is based on the realized range that can be computed from daily high and low price. It has recently been shown to be relatively robust to microstructure noise (Alizadeh et al. (2002)). The Rogers and Satchell (1991) estimator is an enhancement of the Parkinson estimator that makes additional use of the opening and closing prices. Rogers and Satchell (1991) show that their estimator is unbiased for the volatility parameter of a Brownian motion plus drift, whereas the Parkinson estimator is biased in that case. Formally, the Rogers and Satchell volatility estimator can be computed as

$$V_{it_j} = (\ln P^H_{it_j} - \ln P^C_{it_j})(\ln P^H_{it_j} - \ln P^O_{it_j}) + (\ln P^L_{it_j} - \ln P^C_{it_j})(\ln P^L_{it_j} - \ln P^O_{it_j}), \quad (2.1)$$

where $V_{it_j}$ denotes volatility of stock $i$ on day $j$ within week $t$, and $P^O, P^C, P^H, P^L$ are daily opening, closing, high and low prices that are obtained from Datastream. Total volatility increased dramatically during the financial crisis in the latter half of 2008 (Figure 2.3). Figure 2.4 shows total volatility for the FTSE 100 index jointly with entry dates of new venues and latency upgrades at the LSE. Casual inspection suggests that total volatility declined when Turquoise and BATS entered the market. However, this conclusion would be premature because many other events took place at the same time, most importantly, the global financial crisis.

[8] While the available measures of market quality are positive, we wish to emphasize that market quality is a normative concept. Translating positive measures of market quality into welfare is difficult and subject to much controversy (Hart and Kreps (1986), Stein (1987)).

[9] Under MiFID, investment firms are required to seek best execution for their clients, cp. Appendix B.1 for details.

[10] There is a vast econometric literature on volatility measurement and modelling that is summarized by Andersen et al. (2010).

We also decompose total volatility into temporary and permanent volatility. Permanent volatility relates to the underlying uncertainty about the future payoff stream for the asset in question. If new information about future payoffs arrives and that is suddenly impacted in prices, the price series would appear to be volatile, but this is the type of volatility that reflects the true valuation purpose of the stock market. On the other hand, volatility that is unrelated to fundamental information and that is caused by the interactions of traders over- and under-reacting to perceived events is thought of as temporary volatility. [11]To decompose total volatility into a temporary and permanent component, we assume that permanent volatility can be approximated by a smooth time trend. For each stock, temporary volatility is defined as the residuals from the nonparametric regression of total volatility on (rescaled) time (this is effectively a moving average over 1 quarter with declining weights). This approach has been used previously by e.g. Engle and Rangel (2008). The evolution of temporary volatility is shown in the upper right panel of Figure 2.3.

**Liquidity.** Liquidity is a fundamental property of a well-functioning market, and lack of liquidity is generally at the heart of many financial crises and disasters. In practice, researchers and practitioners rely on a variety of measures to capture liquidity. High frequency measures include quoted bid-ask spreads (tightness), the number of orders resting on the order book (depth) and the price impact of trades (resilience). These order book measures may not provide a complete picture since trades may not take place at quoted prices, and so empirical work considers additional measures that take account of both the order book and the transaction record. These include the so-called effective spreads and quoted spreads, which are now widely accepted and used measures of actual liquidity. Another difficulty is that liquidity suppliers often post limit orders on multiple venues but cancel the additional liquidity after the trade is executed on one venue (Kervel and Vincent (2015)). Therefore, global depth measures that aggregate quotes across different venues may overstate liquidity. On the other hand, the presence of "iceberg orders" and dark pools suggest that there is substantial hidden liquidity.

Since we do not have access to order book data, our main measure of liquidity is the percentage bid-ask spread.[12] The quoted bid ask spread for stock $i$ on day $t_j$

[11] A good example is the "hash crash" of April 24, 2013 when the Dow Jones index dropped by nearly 2% very rapidly due apparently to announcements emanating from credible twitter accounts (that had been hacked into) that there had been an explosion at the White House. It subsequently recovered all the losses when it became clear that no such explosion had occurred. See http://blueandgreentomorrow.com/2013/04/24/twitter-hoax-wipes-200bn-off-dow-jones-for-five-minutes/, accessed on June 20, 2013

[12] Mizen (2010) documents that trends in quoted bid-ask spreads are similar to trends in effective bid-ask spreads.

is defined as

$$BA_{it_j} = \frac{P^A_{it_j} - P^B_{it_j}}{\frac{1}{2}(P^A_{it_j} + P^B_{it_j})}, \tag{2.2}$$

where daily ask prices $P^A$ and bid prices $P^B$ are obtained from Datastream. $P^A$ and $P^B$ are measured by the last bid and ask prices before the market closes for London stock exchange at 16:35. The time series of weekly bid-ask spreads is reported in the bottom left panel of Figure 2.3. Inspection of Figure 2.4 seems to suggest that bid-ask spreads declined at the entry of Chi-X but this decline can also attributed to the introduction of Trade Elect 1 at the LSE one day before. Trade Elect 1 achieved a significant reduction of system latency at the LSE.

**Volume.** Volume of trading is a measure of participation, and is of concern to regulators (Foresight (2012)). The volume of trading has increased over the longer term, but the last decade has seen less sustained trend increases, which has generated concern amongst those whose business model depends on this. Some have also argued that computer based trading has led to much smaller holding times of stocks and higher turnover and that this would reflect a deepening of the intermediation chain rather than real benefits to investors.

We investigate both global volume and volume at the LSE. Global volume is defined as the number of shares traded at all venues and volume at the LSE is the number of shares traded at the LSE, scaled by the number of shares outstanding. The volume data is obtained from Fidessa. Towards the end of the sample period, global and LSE volume diverge, as alternative venues gain market share (Figures 3 and 4).

## 2.4 Econometric methodology

Figure 2.3 shows the time series of market quality measures for the FTSE 100 and FTSE 250 indices. All measures clearly show the effect of the global financial crisis that was associated with an increase in total volatility, temporary volatility and bid-ask spreads as well as a fall in traded volumes in the early part of the sample that was followed by reversals (except for volume). As we saw in Figure 1, average fragmentation levels increased for most of the sample. If there were a simple linear relationship between fragmentation and market quality then we would have extrapolated continually deteriorating market quality levels until almost the end of the sample. We next turn to the econometric methods that we will use to exploit the cross-sectional and time series variation in fragmentation and market quality to measure the relationship more reliably.

We extend the CCE estimator of Pesaran (2006) in three ways. First, we allow for some nonlinearity, allowing fragmentation to affect the response variable in a quadratic fashion. This functional form was also adopted in the De Jong et al. (2015) study. Second, we use quantile regression methods based on conditional quantile restrictions rather than the conditional mean restrictions adopted previously.[13] This method is valid under weaker moment conditions and is robust to outliers. A quantile CCE estimator for homogeneous panels is also considered in Harding and Lamarche (2013). Third, we also model the conditional vaiability of market quality using the same type of regression model; we apply the median regression method for estimation based on the squared residuals from the median specification or on the conditional interquartile range. This allows us to analyze not just the average effect of fragmentation on market quality but also the variability of that effect.

### 2.4.1 A model for heterogeneous panel data with common factors

We observe a sample of panel data $\{(Y_{it}, X_{it}, Z_{it}, d_t) : i = 1, \ldots, n,\ t = 1, \ldots, T\}$, where $i$ denotes the $i$-th stock and $t$ is the time point of observation. In our data, $Y_{it}$ denotes market quality and $X_{it}$ is a measure of fragmentation, while $Z_{it}$ is a vector of firm specific control variables such as market capitalization and $d_t$ are observable common factors as for example VIX or the lagged index return. We assume that the data come from the model

$$Y_{it} = \alpha_i + \beta_{1i} X_{it} + \beta_{2i} X_{it}^2 + \beta_{3i}^\intercal Z_{it} + \delta_i^\intercal d_t + \kappa_i^\intercal f_t + \varepsilon_{it}, \tag{2.3}$$

where $f_t \in \mathbb{R}^k$ denotes the unobserved common factor or factors. We allow for a nonlinear effect of the fragmentation variable on the outcome variable by including the quadratic term. We assume that the regression error term satisfies the conditional quantile restrictions

$$\mathrm{Q}_\tau(\varepsilon_{it} | X_{it}, Z_{it}, d_t, f_t) = 0 \tag{2.4}$$

but is allowed to be serially correlated or weakly cross-sectionally correlated. The regressors $W_{it} = (X_{it}, Z_{it}^\intercal)^\intercal$ are assumed to have the factor structure

$$W_{it} = a_i + D_i d_t + K_i f_t + u_{it}, \tag{2.5}$$

where $D_i$ and $K_i$ are matrices of factor loadings. The error term $u_{it}$ is assumed to satisfy $Eu_{it} = 0$ for all $t$, but is also allowed to be serially correlated or weakly cross-sectionally correlated, see Assumptions 1-2 in Pesaran (2006). The econometric model (2.3)-(2.5) also allows for certain types of "endogeneity" between

[13] We provide a justification of this method in Appendix B.4.

the covariates and the outcome variable represented by the unobserved factors $f_t$.[14] The model is very general and contains many homogenous and heterogeneous panel data models as a special case.

We adopt the random coefficient specification for the individual parameters, that is, $\beta_i = (\beta_{1i}, \beta_{2i}, \beta_{3i}^\intercal)^\intercal$ are i.i.d. across $i$ and

$$\beta_i = \beta + v_i, \quad v_i \sim IID(0, \Sigma_v), \tag{2.6}$$

where the individual deviations $v_i$ are distributed independently of $\epsilon_{jt}, X_{jt}, Z_{jt}$ and $d_t$ for all $i, j, t$.

To estimate the model (2.3)-(2.5), we extend Pesaran (2006) CCE mean group estimator to quantile regression. Taking cross-sectional averages of (2.5), we obtain (under the assumption that $u_{it}$ has weak cross-sectional dependence and some finite higher order moments)

$$\overline{W}_t = \bar{a} + \overline{D} d_t + \overline{K} f_t + O_p(n^{-1/2}). \tag{2.7}$$

Equation (2.7) suggests that we can approximate the unknown factor $f_t$ with a linear combination of $d_t$ and the cross-sectional average of $X_{it}$.[15] In contrast to Pesaran (2006), our version of the CCE estimator does not include the cross-sectional average of $Y$. One reason for this is that because of the quadratic functional form, $\overline{Y}_t$ would be a quadratic function of $f_t$, and so would introduce a bias. Instead, we add the Chicago Board Options Exchange Market Volatility Index (VIX) to the specification. Because of the high correlation between VIX and cross-sectional averages of market quality, we expect that VIX is a good and predetermined proxy for cross-sectional averages of market quality in our regressions.

But because only cross-sectional averages of the regressors are used to approximate the common factors, our framework requires that the regressors are driven by the same set of unobserved common factors $f_t$ as the dependent variable.

The effect of fragmentation on market quality can be obtained by performing (for each $i$) a time series quantile regression estimation of (2.3) replacing $f_t$ by $\overline{W}_t$. Specifically, let $\widehat{\theta}_i$ minimize the objective functions

$$\widehat{Q}_{i\tau T}(\theta_i) = \sum_{t=1}^{T} \rho_\tau (Y_{it} - \pi_i - \beta_{i1} X_{it} - \beta_{i2} X_{it}^2 - \beta_{i3}^\intercal Z_{it} - \gamma_i^\intercal d_t - \xi_i^\intercal \overline{W}_t) \tag{2.8}$$

with respect to $\theta_i$, where $\theta_i = (\pi_i, \beta_{i1}, \beta_{i2}, \beta_{i3}^\intercal, \gamma_i^\intercal, \xi_i^\intercal)^\intercal$ and $\rho_\tau(x) = x(\tau - 1(x < 0))$, see Koenker (2005). Then $\widehat{\beta}_i$ are the estimators of the corresponding parameters of

[14] However, the CCE method cannot address simultaneity of Y and X at the individual level due to time varying but firm-specific determinants.

[15] If $f_t$ is a vector, i.e., there are multiple factors, then we must form multiple averages (portfolios). Instead of the equally weighted average in (2.7), we can also use an average that is e.g. weighted by market capitalization.

interest.

At any quantile, the quantile mean group estimator (QCCE) $\widehat{\beta} = n^{-1}\sum_{i=1}^{n}\widehat{\beta}_i$ is defined as the cross-sectional average of the individual quantile estimates $\widehat{\beta}_i = (\widehat{\beta}_{1i}, \widehat{\beta}_{2i}, \widehat{\beta}_{3i}^{\intercal})^{\intercal}$. This measures the average effect. Some idea of the heterogeneity can be obtained by looking at the standard deviations of the individual effects. Following similar arguments as in Pesaran (2006), (as $n \to \infty$) it follows that

$$\sqrt{n}(\widehat{\beta} - \beta) \Longrightarrow N(0, \Sigma), \tag{2.9}$$

where the covariance matrix $\Sigma$ can be estimated by

$$\widehat{\Sigma} = \frac{1}{n-1}\sum_{i=1}^{n}(\widehat{\beta}_i - \widehat{\beta})(\widehat{\beta}_i - \widehat{\beta})^{\intercal}. \tag{2.10}$$

The regression model above concentrates on the average effect, or the effect in normal times. We are also interested in the effect of fragmentation on the variability of market quality. One approach to address this issue is to investigate the conditional variance of market quality. We adopt a symmetrical specification whereby

$$\mathrm{var}(Y_{it}|X_{it}, Z_{it}, d_t, f_t) = a_i + b_{1i}X_{it} + b_{2i}X_{it}^2 + b_{3i}^{\intercal}Z_{it} + w_i^{\intercal}d_t + q_i^{\intercal}f_t, \tag{2.11}$$

where the parameters $b_i = (b_{1i}, b_{2i}, b_{3i}^{\intercal})^{\intercal}$ have a random coefficient specification like (2.6). We estimate this by median regression of the squared residuals $\widehat{\epsilon}_{it}^2$ from (2.3)-(2.5) on $X_{it}, X_{it}^2, Z_{it}, d_t, \overline{W}_t$. We argue in Appendix B.4 that, under suitable regularity conditions, (2.9) holds in this case with a covariance matrix $\Sigma$ (corresponding to the covariance matrix of the parameters of the variance equation).

As an alternative specification for the variability of market quality, we assume that the conditional interquartile range of market quality satisfies

$$q_{0.75}(Y_{it}|X_{it}, Z_{it}, d_t, f_t) - q_{0.25}(Y_{it}|X_{it}, Z_{it}, d_t, f_t) = a_i + b_{1i}X_{it} + b_{2i}X_{it}^2 + b_{3i}^{\intercal}Z_{it} + w_i^{\intercal}d_t + q_i^{\intercal}f_t, \tag{2.12}$$

where $q_\tau(Y_{it}|X_{it}, Z_{it}, d_t, f_t)$ denotes the conditional $\tau$ quantile. (2.12) is estimated by median regression of the conditional interquartile range on $X_{it}, X_{it}^2, Z_{it}, d_t, \overline{W}_t$.

### 2.4.2 Parameter of interest

Motivated by the large increase in market fragmentation over the sample period, we are interested in measuring the market quality at different levels of competition, holding everything else constant. In particular, we would like to compare monopoly with perfect competition. In our data, the maximum number of trading venues is 24 and were trading to be equally allocated to these venues, we might achieve (fragmentation) $X = 0.96$. In fact, the maximum level reached by $X$ is some way below that.

The parameter of interest in our study is the difference of average market quality between a high (H) and low (L) degree of fragmentation or dark trading normalized by $H - L$. We therefore obtain the measure

$$\Delta_X = \frac{E_{X=H}Y - E_{X=L}Y}{H - L} = \beta_1 + \beta_2(H + L), \tag{2.13}$$

where the coefficients are estimated by the QCCE method. For comparison, we also report the marginal effect $\beta_1 + 2\overline{X}\beta_2$. We estimate these parameters from the conditional variance and interquartile range specifications, too, in which case it is to be interpreted as measuring differences in variability between the two market structures. Standard errors can be obtained from the joint asymptotic distribution of the parameter estimates given above.[16]

## 2.5 Results

Before reporting our regression results, we investigate a few characteristics of our dataset in more detail.[17] The particular characteristics we are interested in are cross-sectional dependence and unit roots. The median value of the cross-sectional correlation for different measures of market quality ranges from 0.21 to 0.57 which points to unobserved shocks that are common to many firms. We also carried out a principal-component analysis to investigate if the regressors and the dependent variable are driven by the same set of common shocks. In the regressions where we use fragmentation and market capitalization as regressors, more than 80% of the variance in the regressors and the dependent variable is explained by the first two components (on average across firms), providing evidence for common factors in the data. For the specifications including visible fragmentation, dark trading and market capitalization as regressors, more than 90% of the variation is explained by the first 3 factors. The econometric model we use can control for these common shocks.

We also investigated stationarity of the key variables as this can impact statistical performance, although with our large cross-section, we are less concerned about this.[18] The results from augmented Dickey Fuller tests indicate little support for a unit root in fragmentation or market quality.[19] The average value of fragmentation

[16]An alternative way of comparing the outcomes under monopoly and competition is to compare the marginal distributions of market quality by means of stochastic dominance tests. We report these results in Appendix B.5.

[17]For our empirical analysis, we eliminate all firms with less than 30 observations and all firms where the fraction of observations with zero fragmentation exceeds 1/4. That leaves us with 341 firms for overall fragmentation and 236 firms for visible fragmentation.

[18]Formally, Kapetanios et al. (2011) have shown that the CCE estimator remains consistent if the unobserved common factors follow unit root processes.

[19]The test results are available upon request.

does trend over the period of our study but it has levelled off towards the end and the type of nonstationarity present is not well represented by a global stochastic trend.

### 2.5.1 The effect of total fragmentation, visible fragmentation and dark trading on the level of market quality

Table 2.1 reports QCCE mean group coefficients together with our parameter of interest $\Delta_{Frag}$. $\Delta_{Frag}$ is defined as the difference in market quality between a low and high level of fragmentation evaluated at the minimum and maximum level of fragmentation (equation (2.13)). For comparison, we also report marginal effects, which tend to agree with $\Delta_{Frag}$ in most specifications. As observable common factors, we include VIX, the lagged index return, and a dummy variable that captures the decline in trading activity around Christmas and New Year.[20]

Inspecting $\Delta_{Frag}$, we find that a fragmented market is associated with higher global volume but lower volume at the LSE when compared to a monopoly. These effects are uniform across different quartiles (Table 2.1b)). The increase in global volume in a fragmented market place is consistent with the theoretical prediction in Biais et al. (2000)).

We also find that temporary volatility is lower in a competitive market which is in contrast with what O'Hara and Ye (2011) document using US data for 2008. O'Hara and Ye (2011) also find that fragmentation reduces bid-ask spreads while there is no significant effect in our sample. But O'Hara and Ye (2011) measure market quality globally (using the NMS consolidated order book and trade price), while our measures are local with the exception of global volume.

It is also interesting to split overall fragmentation into visible fragmentation and dark trading where we define dark trading as the sum of volume traded at regulated dark pools, OTC venues and SI (Table 2.2). When measured by $\Delta_{Vis.frag.}$, we find that visible fragmentation reduces temporary volatility and lowers trading volume. These effects are larger in absolute value in the third quartile of the conditional distribution (Table 2.2b)).

In addition, a market with a high degree of visible fragmentation has larger bid-ask spreads at the LSE when compared to a monopoly, albeit that result is only statistically significant at 10%. De Jong et al. (2015) also find that visible fragmentation has a negative effect on liquidity at the traditional exchange. The finding that visible fragmentation may harm local liquidity is also supported by survey evidence: According to Foresight (2012), institutional buy-side investors believe that it is becoming increasingly difficult to access liquidity and that this

[20]The coefficients on the observed common factors and on the cross-sectional averages do not have a structural interpretation because they are a combination of structural coefficients, cf. Section 2.4.1.

is partly due to: its fragmentation on different trading venues, the growth of "dark" liquidity, and to the activities of HFT. To mitigate these adverse effects on liquidity, investors could employ SORT that create a virtually integrated market place. However, the survey reports buy-side concerns that these solutions are too expensive for many investors. In contrast to this evidence, Gresse (2011) finds that visible fragmentation improves local liquidity.

Turning to dark trading, our results suggest that dark trading reduces volatility in particular for firms in the first and second quartile of the conditional volatility distribution (Table 2.2). Dark trading also increases volume while it does not have a significant effect on bid-ask spreads. In comparison, Gresse (2011) also does not find a significant effect of dark trading on liquidity while De Jong et al. (2015) find that dark trading has a detrimental effect on liquidity.

### 2.5.2 Turning points

In addition to investigating the difference between perfect competition and a monopolistic market, it is also interesting to assess the transition between these extremes. Figure 2.5 illustrates the estimated relationship between market quality on the one hand and overall fragmentation, visible fragmentation and dark trading on the other. We find that the transition between monopoly and competition is non-monotonic for overall and visible fragmentation and takes the form of an inverted U shape. The maximum occurs at a level of visible fragmentation of about 0.2, 0.3 and 0.5 for global volume, total volatility and bid-ask spreads, respectively. That is, at low levels of fragmentation, fragmentation of order flow improves market quality but there is a turning point after which fragmentation leads to deteriorating market quality. For temporary volatility and LSE volume, there is no interior optimum on $[0, 1]$.

Securities Exchange Commission (2013) has hypothesized that the turning point may depend on the market capitalization of a stock. For each individual stock, Figure 2.6 plots the maximal level of fragmentation against the time series average of market capitalization.[21] We find that there is positive but weak relationship between the maximal level of fragmentation and market capitalization that is statistically significant with the exception of temporary volatility.[22]

[21] We restrict attention to interior maxima within $[0, 1]$.
[22] These results are qualitatively identical for visible fragmentation.

### 2.5.3 The effect of total fragmentation, visible fragmentation and dark trading on the variability of market quality

In this section, we investigate whether overall fragmentation, visible fragmentation and dark trading have led to an increase in the variability of market quality. For example, Madhavan (2012) finds that higher fragmentation prior to the Flash Crash is associated with larger drawdowns during the Flash Crash. In addition, fragmented equity markets have been a seedbed for HFT that are not obliged to provide liquidity in times of market turmoil. This development can lead to "periodic illiquidity" as for example, during the Flash Crash (Foresight (2012)).

When estimating the conditional variance specification (equation (2.11)), we find that at the median, $\Delta_{Frag.}$ is not statistically significant but there is variation across quartiles (Table 2.3): The variability of volatility is lower in a fragmented market for firms in the third quartile of the conditional distribution. Fragmentation increases the variability of bid ask spreads at the first quartile of the distribution but this result is only marginally significant. There is also a decline in the variability of LSE volume for firms in both the first and third quartile.

Table 2.4 distinguishes between visible fragmentation and dark trading. The effect of visible fragmentation on the variability of volatility are similar to those of overall fragmentation. But in contrast to overall fragmentation, visible fragmentation increases the variability of LSE volume in the first quartile. Dark trading increases the variability of both volatility and volumes but the latter effect is only significant at the first quartile.

Table 2.5 reports the results when the variability of market quality is measured by the conditional interquartile range of volatility (equation (2.12)). Overall, the results are similar: Visible fragmentation reduces the variability of volatility, while dark trading has the opposite effects. Also, dark trading increases the variability of LSE volume.

But there are also some differences between these alternative variability measures: The positive effect of overall and visible fragmentation on the variability in bid-ask spreads is more significant for the inter-quartile range measure of variability when compared to the residual measure. In contrast to the latter, visible fragmentation has no significant effect on the variability of LSE volume.

### 2.5.4 Robustness

In Appendix B.5, we assess the robustness of our results to: (i) alternative market quality measures, (ii) splitting our sample into FTSE 100 and FTSE 250 firms and (iii) different estimation methods. Our finding that visible fragmentation and dark trading have a negative effect on total and temporary volatility is robust to using alternative measures of volatility such as Parkinson or within-day volatility. If we measure market quality by the Amihud (2002) illiquidity measure, we find that a higher degree of overall or visible fragmentation is associated with less liquid markets, and that dark trading is found to improve liquidity. For efficiency, we cannot find significant effects.

When comparing the effect of market fragmentation on market quality for FTSE 100 and FTSE 250 firms, interesting differences emerge: The negative effect of dark trading on volatility is only observed for FTSE 250 firms. That effect is even positive for FTSE 100 firms. But in contrast with FTSE 250 firms, visible fragmentation is associated with lower volatility for FTSE 100 firms.

Finally, we re-estimate our results using a heterogeneous panel data model without common factors. This model can be obtained as a special case of model (2.3)-(2.5) where $f_t$ is a vector of ones and there are no observed common factors $d_t$. A version of this model with homogenous coefficients has been used in related work by Gresse (2011), among others. However, that model cannot account for unobserved, common shocks in the data and gives inconsistent results in the presence of common shocks that are correlated with the regressors (Pesaran (2006)). We report in Appendix B.5 that omitting observed and unobserved common factors leads to results that differ in magnitude and statistical significance with the exception of LSE volume. However, the large value in our measure of cross-sectional dependence (CSD) indicates that this model is misspecified because unobserved common shocks such as changes in trading technology or HFT are omitted that are likely to affect both market quality and fragmentation.

## 2.6 Conclusions

After the introduction of MiFID in 2007, the equity market structure in Europe underwent a fundamental change as newly established venues such as Chi-X started to compete with traditional exchanges for order flow. This change in market structure has been a seedbed for HFT, which has benefited from the competition between venues through the types of orders permitted, smaller tick sizes, latency and other system improvements, as well as lower fees and, in particular, the so-called maker-taker rebates.

Against these diverse and complex developments, identifying the effect of fragmentation on market quality is difficult. To achieve this, we use a version of Pesaran (2006) common correlated effects (CCE) estimator that can account for unobserved factors such as the global financial crisis or HFT. Compared to Pesaran (2006), our QCCE mean group estimator is based on individual quantile regressions that enable us to characterize the whole conditional distribution of the dependent variable rather than just its conditional mean. This estimator is suitable for heterogeneous panel data that are subject to both common shocks and outliers in the dependent variable.

We applied our estimator to a novel dataset that contains weekly measures of market quality and fragmentation for the individual FTSE 100 and 250 firms. We decompose the effect of overall fragmentation into visible fragmentation and dark trading, and assess their effects on both the level and the variability of market quality.

We find that fragmentation and dark trading lower volatility. A more fragmented market is also associated with less variability in volatility in particular at the upper quantiles of the conditional distribution. But dark trading increases the variability of trading volumes and variability which gives rise to some concern.

**Table 2.1:** The effect of fragmentation on the level of market quality

*a) Median regression*

| | Total volatility | Temp. volatility | BA spreads | Global volume | LSE volume |
|---|---|---|---|---|---|
| Constant | -7.745 | -10.511 | 4.468 | 1.713 | 2.365 |
| | (-9.97) | (-17.162) | (5.803) | (2.552) | (3.497) |
| Frag. | 0.45 | -0.856 | 0.195 | 0.064 | 0.413 |
| | (0.805) | (-1.906) | (0.726) | (0.22) | (1.338) |
| Frag. sq. | -0.719 | 0.618 | -0.217 | 0.122 | -1.662 |
| | (-1.619) | (1.694) | (-0.933) | (0.426) | (-5.752) |
| Market cap. | -0.475 | -0.27 | -0.343 | -0.214 | -0.236 |
| | (-6.372) | (-5.767) | (-4.951) | (-3.172) | (-3.492) |
| ME (frag.) | -0.367 | -0.154 | -0.051 | 0.202 | -1.475 |
| | (-3.432) | (-1.823) | (-0.782) | (2.408) | (-18.03) |
| $\Delta_{Frag.}(0.5)$ | -0.15 | -0.341 | 0.014 | 0.166 | -0.973 |
| | (-0.735) | (-2.139) | (0.154) | (1.918) | (-10.108) |
| Adjusted $R^2$ | 0.732 | 0.111 | 0.775 | 0.78 | 0.758 |
| CSD | 0.033 | 0.025 | 0.011 | 0.035 | 0.038 |

*b) Difference between monopoly and competition at* $\tau \in \{0.25, 0.75\}$

| | Total volatility | Temp. volatility | BA spreads | Global volume | LSE volume |
|---|---|---|---|---|---|
| $\Delta_{Frag.}(0.25)$ | -0.219 | -0.356 | -0.067 | 0.14 | -0.944 |
| | (-1.208) | (-2.255) | (-0.818) | (1.677) | (-8.988) |
| $\Delta_{Frag.}(0.75)$ | -0.23 | -0.406 | 0.128 | 0.137 | -0.986 |
| | (-0.982) | (-2.501) | (0.876) | (1.264) | (-8.161) |

*Notes:* Coefficients are quantile CCE mean group estimates. t-statistics are shown in parenthesis. Market capitalization and dependent variables (except of temporary volatility) are in logs. Lagged index return, VIX and Christmas and New Year effects are included as observable common factors. $\Delta_{Frag.}(\tau)$ is defined as $\widehat{\beta}_1(\tau) + \widehat{\beta}_2(\tau)(H + L)$ and evaluated at $H = \max(Frag.)$ and $L = \min(Frag.)$. ME denotes marginal effects. The adjusted $R^2$ is the $R^2$ calculated from pooling the individual total and residual sums of squares, adjusted for the number of regressors. CSD is the mean of the squared value of the off-diagonal elements in the cross-sectional dependence matrix.

**Table 2.2:** The effects of visible fragmentation and dark trading on the level of market quality

*a) Median regression*

| | Total volatility | Temp. volatility | BA spreads | Global volume | LSE volume |
|---|---|---|---|---|---|
| Constant | -8.475 | -11.295 | 1.28 | 1.189 | 2.333 |
| | (-10.602) | (-18.629) | (1.615) | (1.89) | (2.988) |
| Vis. frag. | 0.817 | -0.564 | 0.436 | 0.158 | -0.151 |
| | (2.663) | (-2.171) | (2.085) | (0.759) | (-0.682) |
| Vis. frag. sq. | -1.429 | 0.317 | -0.425 | -0.451 | -1.199 |
| | (-3.937) | (1.019) | (-1.536) | (-1.728) | (-4.323) |
| Dark | -0.212 | 0.388 | -0.212 | 0.332 | 0.232 |
| | (-0.946) | (1.951) | (-1.068) | (1.673) | (1.11) |
| Dark sq. | 0.041 | -0.704 | 0.177 | 1.724 | 0.986 |
| | (0.178) | (-3.47) | (0.897) | (9.605) | (4.867) |
| Market cap. | -0.399 | -0.288 | -0.32 | -0.243 | -0.293 |
| | (-5.328) | (-5.364) | (-4.851) | (-4.29) | (-4.595) |
| ME (vis. frag) | -0.288 | -0.318 | 0.108 | -0.191 | -1.078 |
| | (-2.511) | (-3.405) | (1.394) | (-2.233) | (-13.056) |
| ME (dark) | -0.175 | -0.246 | -0.052 | 1.886 | 1.121 |
| | (-2.628) | (-4.311) | (-1) | (29.009) | (18.205) |
| $\Delta_{Vis.frag.}(0.5)$ | -0.181 | -0.342 | 0.139 | -0.157 | -0.988 |
| | (-1.523) | (-3.537) | (1.86) | (-1.85) | (-11.891) |
| $\Delta_{Dark}(0.5)$ | -0.171 | -0.315 | -0.035 | 2.055 | 1.217 |
| | (-2.518) | (-5.446) | (-0.689) | (34.419) | (20.626) |
| Adjusted $R^2$ | 0.75 | 0.131 | 0.754 | 0.852 | 0.799 |
| CSD | 0.03 | 0.026 | 0.01 | 0.05 | 0.04 |

*b) Difference between monopoly and competition at* $\tau \in \{0.25, 0.75\}$

| | Total volatility | Temp. volatility | BA spreads | Global volume | LSE volume |
|---|---|---|---|---|---|
| $\Delta_{Vis.frag.}(0.25)$ | 0.01 | -0.263 | 0.081 | -0.034 | -0.917 |
| | (0.09) | (-2.879) | (0.959) | (-0.41) | (-11.698) |
| $\Delta_{Vis.frag.}(0.75)$ | -0.487 | -0.61 | 0.112 | -0.22 | -1.094 |
| | (-3.483) | (-5.432) | (1.309) | (-2.036) | (-10.128) |
| $\Delta_{Dark}(0.25)$ | -0.286 | -0.463 | -0.004 | 2.022 | 0.986 |
| | (-3.735) | (-6.63) | (-0.07) | (32.67) | (16.361) |
| $\Delta_{Dark}(0.75)$ | -0.005 | -0.064 | 0.048 | 2.072 | 1.374 |
| | (-0.061) | (-0.935) | (0.785) | (29.979) | (19.166) |

*Notes:* See Table 2.1 except that $X = \{Vis.frag, Dark\}$.

**Table 2.3:** The effect of fragmentation on the variability of market quality (conditional variance model)

*a) Median regression*

| | Total volatility | Temp. volatility | BA spreads | Global volume | LSE volume |
|---|---|---|---|---|---|
| Constant | -0.536 | -0.198 | 0.28 | 0.275 | 0.498 |
| | (-1.893) | (-0.686) | (1.429) | (1.15) | (2.662) |
| Frag. | -0.029 | -0.064 | -0.037 | -0.215 | -0.128 |
| | (-0.256) | (-0.603) | (-0.463) | (-1.716) | (-1.522) |
| Frag. sq. | 0.06 | 0.071 | 0.041 | 0.189 | 0.115 |
| | (0.565) | (0.762) | (0.548) | (1.73) | (1.455) |
| Market cap. | -0.01 | -0.02 | -0.009 | -0.035 | -0.034 |
| | (-0.477) | (-1.099) | (-0.482) | (-2.302) | (-2.312) |
| ME (frag.) | 0.039 | 0.017 | 0.01 | 0 | 0.003 |
| | (1.287) | (0.639) | (0.45) | (-0.001) | (0.139) |
| $\Delta_{Frag.}(0.5)$ | 0.021 | -0.005 | -0.002 | -0.057 | -0.032 |
| | (0.581) | (-0.128) | (-0.096) | (-1.488) | (-1.178) |
| Adjusted $R^2$ | -0.013 | -0.014 | -0.041 | 0.056 | 0.064 |
| CSD | 0.015 | 0.011 | 0.01 | 0.016 | 0.016 |

*b) Difference between monopoly and competition at $\tau \in \{0.25, 0.75\}$*

| | Total volatility | Temp. volatility | BA spreads | Global volume | LSE volume |
|---|---|---|---|---|---|
| $\Delta_{Frag.}(0.25)$ | 0.028 | 0.021 | 0.03 | 0.011 | -0.03 |
| | (1.464) | (1.429) | (1.861) | (0.737) | (-1.847) |
| $\Delta_{Frag.}(0.75)$ | -0.604 | -0.347 | -0.014 | -0.194 | -0.24 |
| | (-2.28) | (-1.921) | (-0.161) | (-1.17) | (-1.82) |

*Notes:* Dependent variables are squared median regression residuals. Coefficients are quantile CCE mean group estimates. t-statistics are shown in parenthesis. Market capitalization is in logs. Lagged index return, VIX and Christmas and New Year effects are included as observable common factors. $\Delta_{Frag.}(\tau)$ is defined as $\widehat{\beta}_1(\tau) + \widehat{\beta}_2(\tau)(H + L)$ and evaluated at $H = \max(Frag.)$ and $L = \min(Frag.)$. ME denotes marginal effects. The adjusted $R^2$ is the $R^2$ calculated from pooling the individual total and residual sums of squares, adjusted for the number of regressors. CSD is the mean of the squared value of the off-diagonal elements in the cross-sectional dependence matrix.

**Table 2.4:** The effect of visible fragmentation and dark trading on the variability of market quality (conditional variance model)

*a) Median regression*

| | Total volatility | Temp. volatility | BA spreads | Global volume | LSE volume |
|---|---|---|---|---|---|
| Constant | -0.708 | -0.034 | 0.208 | -0.145 | 0.054 |
| | (-2.005) | (-0.111) | (0.917) | (-0.972) | (0.287) |
| Vis. frag. | -0.237 | -0.301 | 0.006 | 0.017 | -0.033 |
| | (-1.745) | (-1.545) | (0.089) | (0.314) | (-0.37) |
| Vis. frag. sq. | 0.261 | 0.326 | 0.016 | 0 | 0.094 |
| | (1.546) | (1.453) | (0.17) | (-0.005) | (0.777) |
| Dark | 0.014 | -0.044 | -0.073 | -0.157 | -0.185 |
| | (0.134) | (-0.471) | (-1.13) | (-1.931) | (-2.551) |
| Dark sq. | 0.084 | 0.1 | 0.072 | 0.133 | 0.197 |
| | (0.885) | (1.112) | (1.106) | (2.267) | (3.262) |
| Market cap. | 0.02 | 0.007 | 0.004 | -0.037 | -0.021 |
| | (1.065) | (0.334) | (0.197) | (-2.752) | (-1.378) |
| ME (Vis. frag) | -0.035 | -0.049 | 0.018 | 0.017 | 0.039 |
| | (-0.917) | (-0.928) | (0.661) | (1.136) | (1.633) |
| ME (Dark) | 0.09 | 0.046 | -0.008 | -0.037 | -0.007 |
| | (2.945) | (1.846) | (-0.453) | (-1.138) | (-0.296) |
| $\Delta_{Vis.frag.}(0.5)$ | -0.055 | -0.073 | 0.017 | 0.017 | 0.032 |
| | (-1.359) | (-1.231) | (0.636) | (1.213) | (1.403) |
| $\Delta_{Dark}(0.5)$ | 0.098 | 0.055 | -0.001 | -0.024 | 0.012 |
| | (3.554) | (2.49) | (-0.064) | (-0.853) | (0.619) |
| Adjusted $R^2$ | -0.011 | -0.02 | -0.028 | 0.03 | 0.021 |
| CSD | 0.013 | 0.011 | 0.01 | 0.022 | 0.018 |

*b) Difference between monopoly and competition at $\tau \in \{0.25, 0.75\}$*

| | Total volatility | Temp. volatility | BA spreads | Global volume | LSE volume |
|---|---|---|---|---|---|
| $\Delta_{Vis.frag.}(0.25)$ | 0.052 | -0.007 | 0.007 | 0.009 | 0.019 |
| | (1.701) | (-0.224) | (0.387) | (1.273) | (2.095) |
| $\Delta_{Vis.frag.}(0.75)$ | -0.614 | -0.244 | 0.201 | -0.169 | -0.162 |
| | (-3.145) | (-1.955) | (1.566) | (-1.324) | (-1.228) |
| $\Delta_{Dark}(0.25)$ | 0.03 | 0.022 | 0.011 | 0.013 | 0.024 |
| | (1.771) | (1.853) | (1.211) | (1.966) | (2.599) |
| $\Delta_{Dark}(0.75)$ | 0.19 | 0.223 | 0.028 | -0.07 | -0.046 |
| | (2.054) | (2.66) | (0.387) | (-1.667) | (-0.687) |

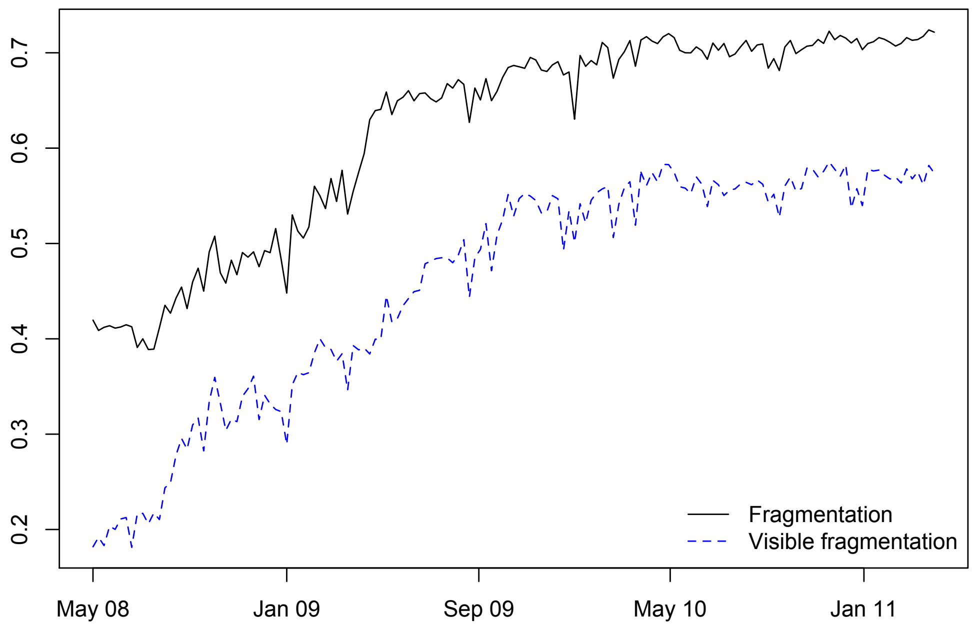
*Notes:* See Table 2.3 except that $X = \{Vis.frag, Dark\}$.

**Table 2.5:** The effects of overall fragmentation, visible fragmentation and dark trading on the variability of market quality (conditional interquartile range model)

| | Total volatility | Temp. volatility | BA spreads | Global volume | LSE volume |
|---|---|---|---|---|---|
| $\Delta_{Frag.}(0.25)$ | -0.021 | 0.021 | 0.214 | -0.001 | -0.038 |
| | (-0.235) | (0.326) | (2.31) | (-0.021) | (-0.418) |
| $\Delta_{Frag.}(0.5)$ | -0.084 | -0.022 | 0.195 | -0.022 | -0.096 |
| | (-0.933) | (-0.347) | (2.111) | (-0.334) | (-1.07) |
| $\Delta_{Frag.}(0.75)$ | -0.09 | -0.041 | 0.179 | -0.058 | -0.106 |
| | (-0.975) | (-0.627) | (1.923) | (-0.873) | (-1.154) |
| $\Delta_{Vis.frag.}(0.25)$ | -0.253 | -0.162 | 0.084 | 0.004 | 0.001 |
| | (-1.67) | (-1.931) | (1.257) | (0.047) | (0.008) |
| $\Delta_{Vis.frag.}(0.5)$ | -0.23 | -0.169 | 0.116 | 0.007 | 0.005 |
| | (-1.524) | (-2.033) | (1.726) | (0.074) | (0.063) |
| $\Delta_{Vis.frag.}(0.75)$ | -0.228 | -0.158 | 0.148 | 0.01 | 0.015 |
| | (-1.501) | (-1.881) | (2.14) | (0.109) | (0.178) |
| $\Delta_{Dark}(0.25)$ | 0.133 | 0.099 | 0.053 | -0.016 | 0.07 |
| | (3.13) | (2.489) | (1.257) | (-0.587) | (2.657) |
| $\Delta_{Dark}(0.5)$ | 0.14 | 0.152 | 0.053 | 0.003 | 0.087 |
| | (3.292) | (3.911) | (1.273) | (0.12) | (3.21) |
| $\Delta_{Dark}(0.75)$ | 0.13 | 0.149 | 0.056 | 0.001 | 0.085 |
| | (3.054) | (3.701) | (1.309) | (0.042) | (3.089) |

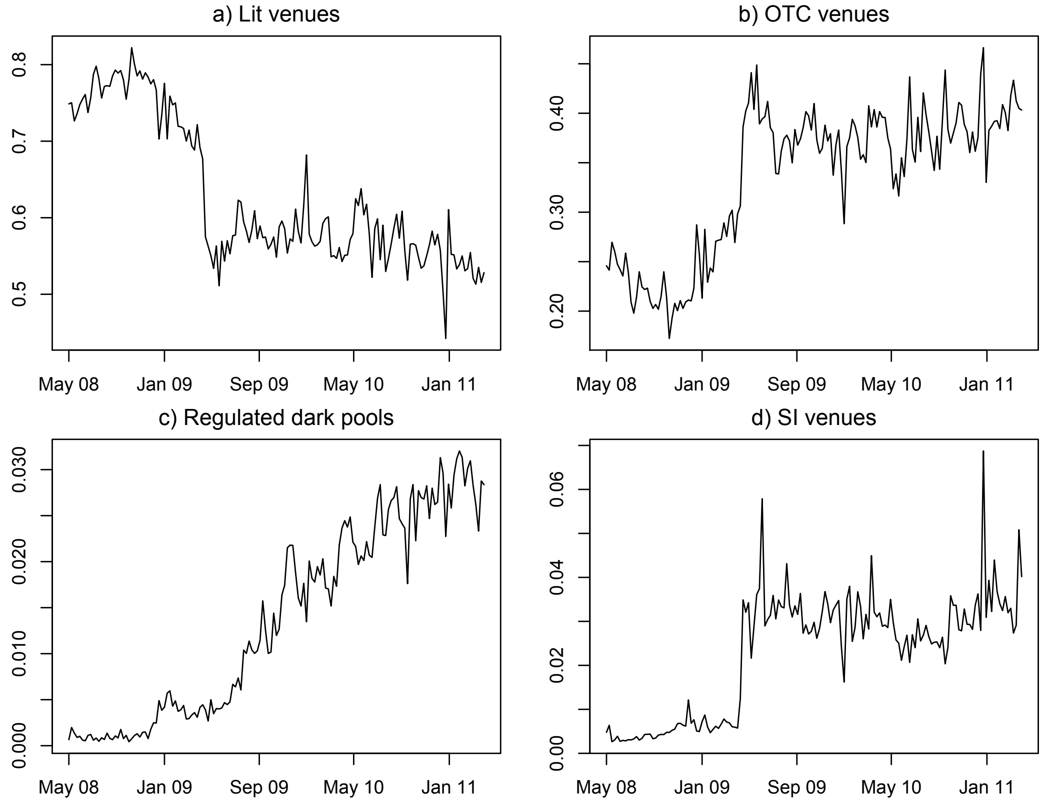
*Notes:* See Table 2.3 except that dependent variables are the conditional interquartile range of market quality.

**Figure 2.1:** Fragmentation and visible fragmentation



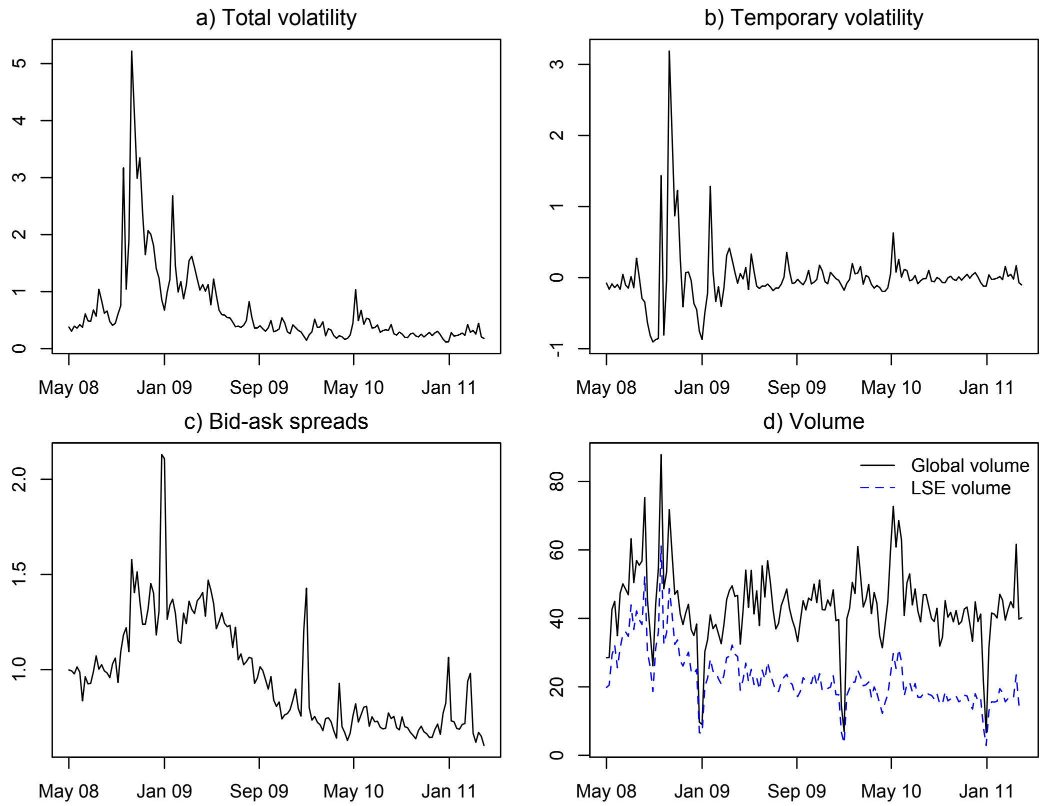

*Notes:* Fragmentation is defined as 1-Herfindahl index and visible fragmentation as 1-visible Herfindahl index. The time series are calculated as averages of the individual series weighted by market capitalization. Data sources: Fidessa, Datastream and own calculations.

**Figure 2.2:** Share of volume traded by venue category



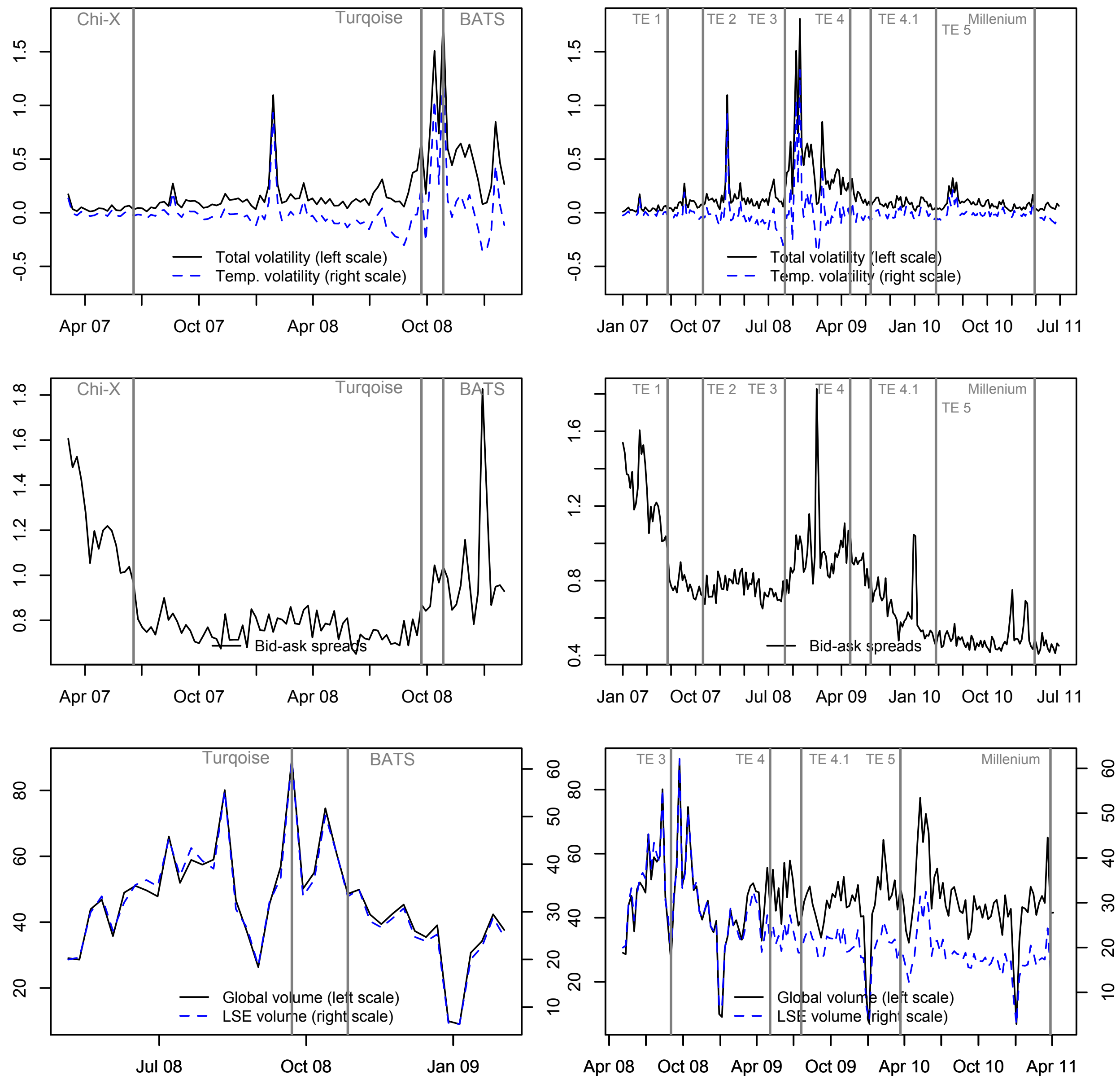

*Notes:* The time series are calculated as averages of the individual series weighted by market capitalization. Data sources: Fidessa, Datastream and own calculations.

**Figure 2.3:** Market quality measures



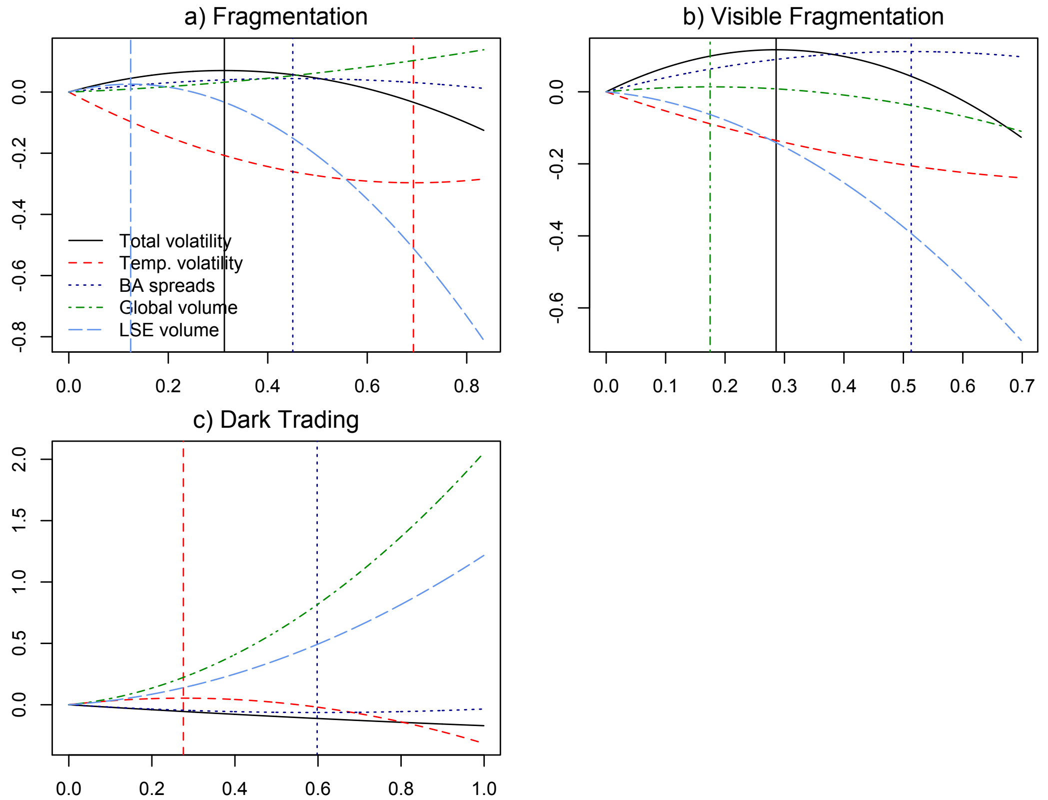

*Notes:* The time series are calculated as averages of the individual series weighted by market capitalization. Bid-ask spreads and volatility are multiplied by 1000. The sharp declines in volume occur during Christmas and New Year. Data sources: Datastream and own calculations.

**Figure 2.4:** Venue entry, latency upgrades at the LSE and market quality for the FTSE 100 index



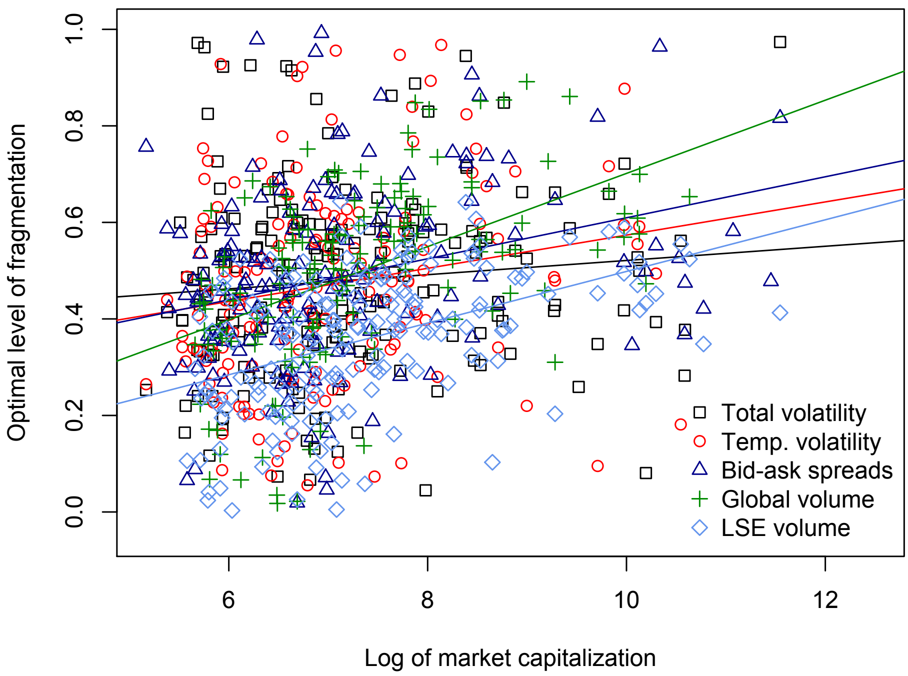

*Notes:* The left panels show market quality measures and venue entry and the right panels show market quality and latency upgrades at the LSE. The time series are calculated as averages of the individual series weighted by market capitalization. Bid-ask spreads and volatility are multiplied by 1000. Series for volume are shorter due to data availability. The sharp declines in volume occur during Christmas and New Year. Data sources: Fidessa, Datastream and own calculations.

**Figure 2.5:** Visible fragmentation, dark trading and market quality




*Notes:* $Y = \widehat{\beta}_1 X + \widehat{\beta}_2 X^2$ is shown where $Y$ is market quality, $X$ is either visible fragmentation, dark trading or OTC trading, and $\widehat{\beta}_j$ are the median CCE estimates from Tables 1 and 2. The vertical lines indicate interior optima.

**Figure 2.6:** The maximal level of fragmentation and market capitalization




*Notes:* The figure plots the optimal level of fragmentation for each individual firm $-\frac{\beta_{1i}}{2\beta_{2i}}$ against the time-series average of the log of market capitalization. Only interior maxima within $[0, 1]$ are shown. OLS regression lines are added.

# Chapter 3

# A Discrete Choice Model For Large Heterogeneous Panels with Interactive Fixed Effects

This paper develops an estimator for heterogeneous panels with discrete outcomes in a setting where the individual units are subject to unobserved common shocks. The proposed estimator belongs to the class of common correlated effects estimators that approximate the unobserved factors with cross-sectional averages. This paper adopts this approach for nonlinear panel data models under the assumption that the unobserved factors are contained in the span of the observed factors and the cross-sectional averages of the regressors. The asymptotic properties of this approach are documented as both the time series and the cross-section tend to infinity. In particular, we show that both the estimators of the individual-specific coefficients and the mean group estimator are consistent and asymptotically normal. The small-sample behavior of the mean group estimator is assessed in a Monte Carlo experiment. The methodology is applied to the question of how funding costs in corporate bond markets affect the conditional probability of issuing a corporate bond.

## 3.1 Introduction

Panel data offer a lot of time series and cross-sectional variation which can facilitate the identification of parameters which are difficult to identify with only time series or cross-sectional data (Mavroeidis et al. (2014)). Recently, researchers have increasingly used panel data to better understand financial and macroeconomic phenomena, such as the transmission of monetary policy (Keys et al. (2014)). In many panel data sets, the popular assumption that individual units are cross-sectionally independent is difficult to maintain. Instead, their behavior is characterized by interdependencies. One source of these dependencies is shocks that are common to all individual units. For example, macroeconomic shocks like financial crises affect household wealth and the balance sheets of firms and financial intermediaries. Both conventional and unconventional monetary policy affects the level of interest rates in the economy and hence consumption decisions of households, investment decisions of firms and the portfolio compositions of financial market participants. Taxes and government subsidies affect the decisions of firms on where to locate a factory, and the decisions of households on whether to take out a mortgage.[1]

Some common shocks are observable while others are not, and the impact of common shocks typically differs across individual units. For example, a bank with a small equity position will reduce its loan supply by more than a well-capitalized bank after a tightening in capital requirements. Andrews (2005) shows that common shocks create problems for inference if data are available for a single cross-sectional unit and the model is estimated by least squares or instrumental variable methods. But the increased availability of panel data where both the time series and cross-sectional dimensions are large offers new opportunities for controlling for these unobserved shocks (Bai (2009), Pesaran (2006)).

This paper contributes to that literature by developing an estimator for large heterogeneous panels with cross-sectional dependence in a framework where outcomes are discrete.[2] The proposed estimator belongs to the class of common correlated effects (CCE) estimators that approximate the unobserved factors with cross-sectional averages (Pesaran (2006)). But this approach is complicated in nonlinear models where the unobserved factors and the cross-sectional averages are linked by an unknown functional form.[3] This paper adopts the CCE estimation methodology to discrete choice models under the assumption that the unobserved factors are contained in the span of the observed factors and the cross-sectional

[1] Andrews (2003) discusses a variety of common shocks including macroeconomic, political, legal, environmental and health shocks.

[2] Discrete choice models are probably the most popular nonlinear panel data models in econometrics. But the methodology developed here is also applicable to other nonlinear panel data models.

[3] One exception is the state-dependent pricing model of Dhyne et al. (2011) where it is possible to solve for the unobserved factors as a function of cross-sectional averages.

averages of the regressors. In situations where the number of unobserved factors is likely to be large relative to the number of regressors, I discuss how instrumental variable methods can be applied. Finally, I sketch a more general approach to adopt the CCE methodology to nonlinear panels that is based on simultaneous sieve M-estimation.

To derive the asymptotic properties of the estimators of the individual-specific coefficients and their mean, I take a perspective where both the time dimension $T$ and the cross-sectional dimension $N$ are large. I first show that the estimator of the individual-specific coefficients is consistent and asymptotically normal. It turns out that this estimator has the same asymptotic distribution as an infeasible estimator that counterfactually assumes that the unobserved factors are known. An important part of the asymptotic theory is uniform consistency of the preliminary estimator. Specifically, the preliminary estimator converges to the true function at the uniform rate $\log(T)/\sqrt{N}$.

Based on the asymptotic properties of the estimators of the individual-specific coefficients, consistency and asymptotic normality results for the mean group estimator are derived. Inference is easy: I show that the asymptotic variance of the mean group estimator can be estimated by the covariance matrix of the individual-specific coefficient estimates. This covariance estimator is identical to the estimator obtained in linear regression models (Pesaran (2006)).

By means of a simulation study, I document that for a wide range of factor structures, the mean group estimator is comparable in terms of RMSE and bias to an infeasible estimator that counterfactually assumes that the common factors are known. In addition, the mean group estimator has good empirical power and size.

I apply the methodology developed in this chapter to the question of how yields in corporate bond markets affect the conditional probability of issuing a bond. I find that conditional probability of issuing a bond is larger in low yield environments for non-financial firms. This question is of policy interest because it sheds light on a particular transmission mechanism of monetary policy: central banks can affect the interest rates that firms face in corporate bond markets by means of conventional and unconventional monetary policy tools. Bond issuance, on the other hand, is often related to corporate investment.

Recently, a growing literature has documented that CCE estimators are consistent and asymptotically normal in a variety of situations, such as quantile regression (Harding and Lamarche (2013)), structural breaks (Baltagi et al. (2016)), and dynamic panels (Chudik and Pesaran (2015a)). But there is not yet a CCE estimator for panels with binary outcomes when both N and T are large which is the contribution of this chapter. Alternative approaches to estimation and inference in nonlinear panel data models with interactive fixed effects are the two-step estimators of Chen (2014) and Chen et al. (2014).[4] In contrast to the estimator proposed here, these estimators are computed in an iterative procedure and assume that the slopes are homogeneous which requires bias correction due to the incidental parameter problem.

The remainder of this chapter is organized as follows. Section 3.2 discusses alternative approaches to model cross-sectional dependence in disturbance terms. The econometric model is presented in Section 3.3 and Section 3.4 develops the estimation methodology. Section 3.5 establishes the asymptotic properties of the estimators of the individual-specific coefficients and the mean group estimator. Section 3.6 reports the results of a Monte Carlo experiment and Section 3.7 applies the methodology to the question of how yields affect the decision to issue a corporate bond. Section 3.8 concludes.

## 3.2 Modeling cross-sectional dependence

A variety of alternative approaches have been proposed to address cross-sectional dependence in disturbance terms. Broadly, they can be classified into two main categories: spatial processes and factor structures. The remainder of this section discusses both approaches in more detail.[5]

Spatial models assume that cross-sectional dependence arises because of interactions among economic agents. Such interactions are predicted by economic theory. For example, when faced with idiosyncratic shocks, rational agents will take out insurance contracts to smooth consumption which makes individual consumption profiles cross-sectionally correlated (Conley (1999)). Other examples include trade or financial spill-over effects.

Conley (1999) observed that the strength of these interactions is related to the economic distance between individual units. Economic distance is determined by a variety of socio-economic characteristics and it can be measured by a distance metric as for example the Euclidean norm. Once a spatial ordering is established, Conley (1999) adopts mixing coefficients from the time-series literature to study the asymptotic properties of GMM estimation.

A parametric alternative to mixing conditions are spatial autoregressive frameworks (SAR). They rely on a weight matrix that summarizes the strength of the cross-sectional correlations. The weight matrix is usually known up to a small number of parameters. However, these methods assume that the econometrician has prior knowledge about how to measure interactions.

Recently, Robinson (2011) has proposed an alternative approach based on linear processes that does not need a measure of cross-sectional distance and includes the SAR model as a special case. This approach was used in e.g. Lee (2012) to

[4] I learned about these papers after completing a first draft of this paper.

[5] See also Chudik and Pesaran (2015b), Bailey et al. (2015) and Lee (2012) for an overview.

derive the asymptotic properties of a series estimator in panels with cross-sectional dependence.

But spatial approaches assume that the cross-sectional dependence is local or weak as defined in Chudik et al. (2011). If the data exhibit strong cross-sectional dependence, factor structures are better suited to model it. The motivation behind factor structures is that there are common but unobserved shocks that can have different impacts on the individual units. These shocks are allowed to be correlated with the regressors which is more general compared to spatial settings. The unobserved factors can be estimated by principal components methods (Coakley et al. (2002), Bai (2009)), or by cross-sectional averages (Pesaran (2006)).

Coakley et al. (2002) develop a two-step principal components estimator where in a first step, the common factors are extracted by computing the principal components from the OLS residuals. In the second step, the regression model augmented with the principal components is estimated. However, as shown by Pesaran (2006), this estimator is not consistent if the unobserved common factors are correlated with the explanatory variables. To overcome this problem, Bai (2009) develops an iterative principal components estimator that alternates the first and second step of Coakley et al. (2002) until convergence. The iterative principal components estimator is consistent even if regressors and unobserved factors are correlated.

The common correlated effects (CCE) estimator of Pesaran (2006) is based on the idea that the unobserved factors can be approximated by cross-sectional averages of the dependent and independent variables. Compared to the principal components estimator, CCE estimators have the advantage that the number of common factors does not need to be known. In Monte Carlo studies, CCE estimators are found to be more efficient and robust when compared to alternative estimators including principle component estimators (Coakley et al. (2006), Chudik et al. (2011)).

A growing literature has documented that CCE estimators are consistent in a variety of situations: Pesaran and Tosetti (2011) combine the factor approach with spatial models by assuming that the disturbances net of the common factors follow a spatial process, see also Chudik et al. (2011). Kapetanios et al. (2011) show that the CCE estimator is consistent even if the unobserved factors are non-stationary. Chudik and Pesaran (2015a) extend the CCE estimator to dynamic panels.[6] Baltagi et al. (2016) develop a CCE estimator for data with structural breaks. Harding and Lamarche (2013) propose a quantile CCE estimator for homogeneous panel data with endogenous regressors, and Boneva et al. (2015) develop a quantile CCE estimator for heterogeneous panels. The contribution of this chapter is to extend the CCE approach to discrete outcomes.

[6]Alternative estimators for dynamic panels with cross-sectional dependence are Moon and Weidner (2015) and Song (2013).

## 3.3 Econometric model

This section describes the econometric framework. I observe a sample of panel data $\{(Y_{it}, X_{it}, d_t) : i = 1, \ldots, n,\ t = 1, \ldots, T\}$, where $i$ denotes the $i$-th unit and $t$ is the time point of observation. To keep the notation simple, I assume that the panel is balanced. The data are assumed to come from the model

$$Y_{it}^* = \alpha_i^\intercal d_t + \beta_i^\intercal X_{it} + e_{it}, \quad i = 1, \ldots, N, t = 1, \ldots, T, \tag{3.1}$$

where $Y_{it}^*$ is a latent variable that is related to the observed response variable $Y_{it}$ via the indicator function $I(.)$,

$$Y_{it} = I(Y_{it}^*). \tag{3.2}$$

That is, $Y_{it}$ is unity if $Y_{it}^* > 0$ and zero otherwise. Alternatively,

$$\Pr(Y_{it} = 1 | X_{it}, d_t, f_t) = 1 - \Phi(-\alpha_i^\intercal d_t - \beta_i^\intercal X_{it} - \kappa_i^\intercal f_t) = \Phi(\alpha_i^\intercal d_t + \beta_i^\intercal X_{it} + \kappa_i^\intercal f_t),$$

where $\Phi\ (.)$ is the standard normal CDF. $X_{it}$ is a $K_x \times 1$ vector of individual-specific regressors that are assumed to be strictly exogenous and stationary and $d_t$ is a $K_d \times 1$ vector of observed common factors that do not vary across individual units. This chapter is concerned with inference for the heterogeneous coefficients $\beta_i$ and their mean. This is complicated by cross-sectional dependence which is modeled by assuming that the disturbances exhibit the factor structure

$$e_{it} = \kappa_i^\intercal f_t + \epsilon_{it}, \tag{3.3}$$

where $f_t$ is a $K_f \times 1$ vector of unobserved common factors and $\kappa_i$ is a $K_f \times 1$ vector of factor loadings. The disturbances $\epsilon_{it}$ are $IID$ conditional on the factors and have a normal distribution with zero mean and unit variance (although our method can be defined for any link function with some regularity conditions). The normalization of the variance is necessary for identification of $\beta_i$.[7]

The panel data model (3.1)-(3.3) contains popular panel data models with additive factor structures as a special case. For example, if $\beta_i$ and $\kappa_i$ are homogeneous across $i$ and $d_t$ only includes a constant, the model reduces to a discrete choice panel model with homogeneous slopes and individual and time fixed effects. As documented in Fernandez-Val and Weidner (2015), this model is subject to the incidental parameter problem (Neyman and Scott (1948)) which results in biased estimates that need to be corrected with jackknife methods, for example. By assuming that the coefficients in the discrete choice panel data model (3.1)-(3.3) are heterogeneous, I can avoid this problem.

In many panel data applications, the unobserved common factors $f_t$ are

[7]Alternative identification assumptions can be made as e.g. $\alpha_{1i} = 1$.

correlated with both the response variable and the regressors, introducing a certain type of endogeneity. To allow for this possibility, the individual-specific regressors are assumed to follow the model

$$X_{it} = A_i^\intercal d_t + K_i^\intercal f_t + u_{it}, \tag{3.4}$$

where $A_i$ is a coefficient matrix of dimension $K_d \times K_x$, $K_i$ is a $K_f \times K_x$ matrix of factor loadings and $u_{it}$ have a zero mean and are $IID$ conditional on the common factors.

To analyze the asymptotic properties of the estimators of the individual-specific coefficients $\widehat{\beta}_i$ and the mean group estimator

$$\widehat{\beta} = \frac{1}{N}\sum_{i=1}^{N}\widehat{\beta}_i. \tag{3.5}$$

I make the following assumptions which are maintained throughout the chapter:

(A1) *Random coefficient model*: the coefficients $\beta_i$ are generated by

$$\beta_i = \beta + \eta_i, \tag{3.6}$$

where $\eta_i \sim IID(0, \Sigma_\eta)$ and is distributed independently of $\kappa_j, K_j, \epsilon_{jt}, u_{jt}, d_t, f_t$ $\forall i, j, t$, $\|\beta\| < C_\beta$, $\|\Sigma_\eta\| < C_\Sigma$ is symmetric and non-negative definite and $\|A\| = \sqrt{\mathrm{tr}(A^\intercal W A)}$ for any matrix A and a symmetric positive definite matrix W.

(A2) *Common factors*: the $(K_f + K_d) \times 1$ vector of common factors $g_t = (f_t^\intercal, d_t^\intercal)^\intercal$ is assumed to be bounded and covariance stationary with absolute summable covariances, and distributed independently of the disturbances $\epsilon_{it}$ and $u_{is}$, $\forall i, t, s$.

(A3) *Factor loadings*: the factor loadings $\kappa_i$ and $K_i$ are IID across $i$, and distributed independently of the disturbances $\epsilon_{jt}$ and $u_{jt}$ and the common factors $f_t$ and $d_t$, for all $i, j, t$ with finite means and variances.

## 3.4 A common correlated effects estimator for discrete choice panels

The econometric model (3.1)-(3.4) depends on the unobserved factors $f_t$ which makes estimation difficult. One approach to control for unobserved factors is to approximate them by cross-sectional averages of $X_{it}$ and $Y_{it}$ (Pesaran (2006)). Section 3.4.1 adopts this approach for discrete choice models under the assumption that the unobserved factors are contained in the span of the observed factors and the cross-sectional averages of the regressors.[8] In situations where the number of unobserved factors is likely to be large relative to the number of regressors, instrumental variable methods can be applied. This is discussed in Section 3.4.2. Section 3.4.3 sketches a more general approach based on simultaneous sieve M-estimation.

### 3.4.1 Approximating the unknown factors by cross-sectional averages of the regressors

In nonlinear panel data models, approximating the unobserved factors by cross-sectional averages of both the regressors and the dependent variable as in Pesaran (2006) is difficult. Instead, I adopt the same approach as in chapter 2 of this thesis and approximate the unobserved factors by cross-sectional averages of the regressors. This approach, however, assumes that the regressors are driven by the same set of factors as the dependent variable.

I start by taking cross-sectional averages of equation (3.4) to obtain

$$\begin{aligned}\overline{X}_t &= \overline{A}^\intercal d_t + \overline{K}^\intercal f_t + \overline{u}_t \\ &= {A_0}^\intercal d_t + {K_0}^\intercal f_t + \overline{u}_t + (\overline{A} - A_0)^\intercal d_t + (\overline{K} - K_0)^\intercal f_t \\ &= {A_0}^\intercal d_t + {K_0}^\intercal f_t + O_p(1/\sqrt{N}).\end{aligned} \tag{3.7}$$

Under the assumption that

$$\text{rank}(\overline{K}) = K_f \leq K_x \; \forall N, \tag{3.8}$$

the unobserved factors can be represented as

$$f_t = (\overline{K}\,\overline{K}^\intercal)^{-1}\overline{K}\,\overline{X}_t - (\overline{K}\,\overline{K}^\intercal)^{-1}\overline{K}\,\overline{A}^\intercal d_t - (\overline{K}\,\overline{K}^\intercal)^{-1}\overline{K}\,\overline{u}_t,$$

where it is possible to show that $\overline{u}_t \xrightarrow{p} 0$ (Pesaran (2006)). Therefore, the unobserved factors can be approximated by a linear combination of cross-sectional averages of the regressors the observed common factors.[9]

### 3.4.2 Instrumental variables

But in situations where the number of unobserved factors is likely to be large relative to the number of regressors, the approach outlined above may not work very well.

[8] In case of microeconometric panels where individual-specific unobserved characteristics like ability are likely to be correlated with the regressors, the indices $t$ and $i$ can be interchanged and time-series averages can be used to approximate the unobserved loadings.

[9] Rather than taking simple averages, any weighted average can be used provided that the weights are "granular" as defined in Pesaran (2006)

This section describes how instrumental variable methods can be applied in these cases.

If an instrument $W_{it}$ for $f_t$ is available that is related to $Y_{it}$ only via the common factors $f_t$ then that instrument can be used to approximate the common factors together with $\overline{X}_t$. Specifically, assume that there is an instrument available that is generated as

$$W_{it} = A_i^W + K_i^{W^\intercal} f_t + w_{it},$$

where $w_{it}$ is an $IID$ disturbance term. The unobserved factors can then be approximated by a linear combination of $d_t$ and $\overline{Z}_t = [\overline{X}_t^\intercal : \overline{W}_t^\intercal]^\intercal$.

### 3.4.3 A more general approach based on sieve estimation

In Section 3.4.1, it was assumed that the unobserved factors are contained in the span of the observed factors and the cross-sectional averages of the regressors. This is more restrictive compared to the CCE approach in Pesaran (2006) that also approximates the unobserved factors with the cross-sectional average of the dependent variable. This section sketches a more general approach for CCE-type estimation in nonlinear panels that also includes the cross-sectional average of the dependent variable. The individual regression models are first averaged to obtain

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N}\Pr(Y_{it}=1|d_t, X_{it}, f_t) &= \frac{1}{N}\sum_{i=1}^{N}\Phi(\alpha_i^\intercal d_t + \beta_i^\intercal X_{it} + \kappa_i^\intercal f_t)\\ &= H_N(d_t, f_t, \mu_t^X),\end{aligned} \tag{3.9}$$

where $N^{-1}\sum_{i=1}^{N}\Pr(Y_{it} = 1|d_t, X_{it}, f_t) = \overline{Y}_t$ because $Y_{it}$ is a discrete random variable. $\mu_t^X$ is a vector with cross-sectional sample moments of $X_{it}$ that completely characterizes the cross-section of $X_{it}$. For simplicity, I assume that $\mu_t^X = \overline{X}_t$ but the approach outlined in this section can be generalized to cases where $\mu_t^X$ also contains higher moments.[10] This assumption is equivalent to assuming that higher moments are time-invariant. Equation (3.9) implies that $f_t$ is given by [11]

$$f_t = G(\overline{Y}_t, \overline{X}_t, d_t), \tag{3.10}$$

where $G(.)$ is an unknown function that can be estimated by series estimation, for example.[12] Series estimation replaces the unknown function $G(.)$ by the first $q$ terms of a sequence of approximating functions $p^q(\overline{Y}_t, \overline{X}_t, d_t)^\intercal\zeta_i$ where $p^q(.) = (p_1(.), \ldots, p_q(.))^\intercal$ and $\zeta_i$ is a fixed vector of parameters. I consider the case where $p_j(.)$ are

[10] $\mu_t^X = \overline{X}_t$ holds exactly if $H_N(.)$ is linear.

[11] A sufficient (but not necessary) condition for uniqueness of $f_t$ is $\frac{\partial H_N}{\partial f} > 0$ or $\frac{\partial H_N}{\partial f} < 0$.

[12] The dependence on N is implicit.

multivariate polynomials of order $j$ which can be motivated by a Taylor expansion.[13]

Newey (1994) develops a two-step series estimator for the case where $f_t$ is known. If $f_t$ is unobserved, the individual-specific coefficients $\beta_i$ can be estimated in one-step by replacing the unobserved factors $f_t$ by the approximating functions $p^q(\overline{Y}_t, \overline{X}_t, d_t)^\intercal \zeta_i$:[14]

$$\Pr(Y_{it=1}|d_t, X_{it}, f_t) \approx \Phi(\alpha_i^\intercal d_t + \beta_i^\intercal X_{it} + \kappa_i'^q(\overline{Y}_t, \overline{X}_t, d_t)^\intercal \zeta_i). \tag{3.11}$$

In general, $q$ is chosen by data-driven procedures such as cross-validation, for example. In same applications, including higher order terms like cross-sectional variances can be desirable. For a general link function $F(.)$, equation (3.11) is a semiparametric regression model that can be estimated by simultaneous sieve M-estimation, see Chen (2007) for a survey.

## 3.5 Asymptotic theory

This section characterizes the asymptotic properties of both the estimators of the individual-specific coefficients and the mean group estimator in discrete choice panels with interactive fixed effects.

**Notation:** the true individual-specific coefficients and their population means are denoted by $\theta_{0i} = (\bar{\alpha}_{0i}^\intercal, \beta_{0i}^\intercal, \bar{\kappa}_{0i}^\intercal)^\intercal$ and $\theta_0 = (\bar{\alpha}_0^\intercal, \beta_0^\intercal, \bar{\kappa}_0^\intercal)^\intercal$ where I use $\bar{}$ to denote the coefficients in the regression model augmented with cross-sectional averages. I define the $K_x \times 1$ vector $\widehat{h}_t \equiv \overline{X}_t = \overline{A}^\intercal d_t + \overline{K}^\intercal f_t + \overline{u}_t$ which can be interpreted as an estimator of the $K_x \times 1$ vector $h_{0t} = A_0{}^\intercal d_t + K_0{}^\intercal f_t$. $\widehat{h}_t$ and $h_{0t}$ can be stacked to form the $TK_x \times 1$ vectors $\widehat{h}(T) = (\overline{X}_{1,1}, \ldots, \overline{X}_{K_x,1}, \overline{X}_{1,2}, \ldots, \overline{X}_{K_x,2}, \ldots, \overline{X}_{1,T}, \ldots, \overline{X}_{K_x,T})$ and $h_0(T) = (h^0_{1,1}, \ldots, h^0_{K_x,1}, h^0_{1,2}, \ldots, h^0_{K_x,2}, h^0_{1,T}, \ldots, h^0_{K_x,T})$, which can be embedded within the sequence space $\mathcal{H}$ whose metric is $d(h,g) = \sup_{i\geq 1}|h_i - g_i|$, in which case I write $h_0 = (h^0_{1,1}, \ldots, h^0_{K_x,1}, h^0_{1,2}, \ldots, h^0_{K_x,2}, h^0_{1,T}, \ldots, h^0_{K_x,T}, 0, \ldots)$ and likewise $\widehat{h}$ (suppressing dependence on $T$), and let $\|h\|_{\mathcal{H}} = d(h,h)$. I use $\Theta$ to denote the finite dimensional parameter set for $\theta_i$ (where the dependence on $i$ is suppressed) and $\mathcal{H}$ for the infinite dimensional parameter set of sequences $\{h_t\}_{t=1}^{\infty}$. C denotes a finite constant.

### 3.5.1 Asymptotics for the estimators of the individual-specific coefficients

This section shows that the estimators of the individual-specific coefficients $\widehat{\theta}_i$ are consistent and have the same asymptotic distribution as an infeasible estimator

[13]To reduce multicolinearity, orthogonal polynomials such as Chebyuchev polynomials can be used instead.

[14]The coefficients $\kappa_i$ and $\zeta_i$ are not individually identified.

that counterfactually assumes that the unobserved common factors $f_t$ are known. Observe that the vector $\widehat{\theta}_i$ contains both the coefficients of interest $\widehat{\beta}_i$ and the auxiliary coefficients on the known factors $d_t$ and the cross-sectional averages $\overline{X}_t$ which play the role of nuisance parameters. For notational simplicity, the asymptotic theory is presented for $\widehat{\theta}_i$ rather than for the parameter of interest $\widehat{\beta}_i$. [15]

The estimator $\widehat{\theta}_i$ is defined to minimize minus the log-likelihood function

$$\begin{aligned}\widehat{Q}_T^i(\theta_i) = Q_T^i(\theta_i, \widehat{h}) &= -\frac{1}{T}\sum_{t=1}^{T}\log f(Y_{it}|X_{it}, \widehat{h}_t, d_t, \theta_i)\\ &= -\frac{1}{T}\sum_{t=1}^{T}[Y_{it}\log F(\overline{\alpha}_i^\intercal d_t + \beta_i^\intercal X_{it} + \overline{\kappa}_i^\intercal \widehat{h}_t)\\ &\quad + (1 - Y_{it})\log(1 - F(\overline{\alpha}_i^\intercal d_t + \beta_i^\intercal X_{it} + \overline{\kappa}_i^\intercal \widehat{h}_t))].\end{aligned}$$

The probability limit of $\widehat{Q}_T^i(\theta_i, h)$ for a given sequence $\{h_t\}$ is defined as

$$Q_0^i(\theta_i, h) = -E\left[\log f(Y_{it}|X_{it}, h_t, d_t, \theta_i)\right].$$

Define likewise the infeasible objective function

$$Q_T^i(\theta_i) = Q_T^i(\theta_i, h_0) = -\frac{1}{T}\sum_{t=1}^{T}\log f(Y_{it}|X_{it}, h_{0t}, d_t, \theta_i). \tag{3.12}$$

An important condition to derive the asymptotic properties of $\widehat{\theta}_i$ is uniform consistency of $\widehat{h}_t$. To obtain the uniform convergence rate of

$$\widehat{h}_t - h_{0t} = \overline{u}_t - (A_0 - \overline{A})^\intercal d_t - (K_0 - \overline{K})^\intercal f_t,$$

I make the following assumptions:[16]

(B1) $E(|u_{it}^j|^k) < C_v$, $k \geq 6$, where $u_{it}^j$ denotes the $j$th element in the $K_x \times 1$ vector $u_{it}$

(B2) $E\|A_i - A_0\|^2 \leq C < \infty$, $E\|K_i - K_0\|^2 \leq C < \infty$

(B3) $T/N \to 0$

Assumption (B1) is required for Bernstein's exponential inequality. Condition (B2) requires that the matrices $A_i$ and $K_i$ have finite variances and is implied by

[15] The coefficients $a_i$ on the observed common factors $d_t$ are not identified in the regression model augmented with cross-sectional averages. But under the assumption that the unobserved factors $f_t$ are orthogonal to the observed factors $d_t$, there are two possible approaches to estimate $\widehat{a}_i$: (i) estimate $X_{it} = A_i^\intercal d_t + \tilde{u}_{it}$, where $\tilde{u}_{it} = K_i^\intercal f_t + u_{it}$; (ii) estimate $\widetilde{Y}_{it} = a_i^\intercal d_t + \widetilde{e}_{it}$ where $\widetilde{Y}_{it} = Y_{it} - \widehat{\beta}X_{it}$.

[16] Random variables are understood as triangular arrays of random variables. This is left implicit to keep the notation simple.

assumptions (A1) and (A3) provided that the parameters in $A_i$ have finite second moments. The following lemma gives an upper bound on the uniform convergence rate of $\widehat{h}_t - h_{0t}$.[17]

**Lemma 3.5.1.** *Suppose that assumptions (A2), (B1)-(B3) hold. Then*

$$\|\widehat{h} - h_0\|_{\mathcal{H}} = O_p\left(\frac{\log T}{\sqrt{N}}\right). \tag{3.13}$$

*Furthermore, for weight sequence $\{\omega_t\}$ with $\sum_{t=1}^{T}\omega_t = 1$ and $\sum_{t=1}^{T}\omega_t^2 \leq C$, I have*

$$\sum_{t=1}^{T}\omega_t\left(\widehat{h}_t - h_{0t}\right) = O_p\left(\frac{1}{\sqrt{N}}\right).$$

An important assumption to obtain the uniform rate in (3.13) is boundedness of the factors that is assumed in (A2). If the factors are not bounded, the penalty term is $T^\delta$ instead of $\log T$ where $\delta$ depends on the number of moments that the observed and unobserved factors $g_t$ possess.

After having established uniform consistency of the preliminary estimator, I next show that the estimators of the individual-specific coefficients is consistent. To this objective, I make the following assumptions:

(C1) The parameter space $\Theta$ is compact and $\theta_{i0} \in \Theta$

(C2) $\widehat{\theta}_i \in \Theta$ and $Q_T^i(\widehat{\theta}_i, \widehat{h}) = \inf_{\theta_i \in \Theta} Q_T^i(\theta_i, \widehat{h})$

(C3) $\|\widehat{h} - h_0\|_{\mathcal{H}} = o_p(1)$

(C4) For all $\delta_i > 0$, there exists $\epsilon(\delta_i)$ such that

$$\inf_{\|\theta_i - \theta_{0i}\| > \delta}\left|Q_0^i(\theta_i, h_0) - Q_0^i(\theta_{0i}, h_0)\right| \geq \epsilon(\delta_i) > 0$$

(C5) $\widetilde{\theta}_i \xrightarrow{p} \theta_{0i}$ for each fixed $i$ where $\widetilde{\theta}_i = \arg\min_{\theta \in \Theta} Q_T^i(\theta, h_0)$ and $Q_T^i(\theta, h_0)$ is defined in (3.12)

(C6) For $\delta_T = o_p(1)$,

$$\sup_{||h - h_0|| \leq \delta_T}\sup_{\theta \in \Theta}\left|Q_T^i(\theta, h) - Q_T^i(\theta, h_0)\right| = o_p(1).$$

Compactness of the parameter space (C1) can be dropped provided that the log-likelihood function is concave. Here, concavity of the log-likelihood follows from concavity of $\log F(v)$ and $1 - \log F(v)$ because $v = \overline{\alpha}_i^\intercal d_t + \beta_i^\intercal X_{it} + \overline{\kappa}_i^\intercal h_t$ is linear in the coefficients given $h_t$. But $\log F(v)$ is concave in $v$ because the first derivative

[17] All proofs are relegated to Appendix C.1.

of $\log F(v)$ with respect to $v$ is monotonically decreasing (Newey and McFadden (1994)). Assumption (C2) defines the estimator and can be weakened to $Q_T^i(\widehat{\theta}_i, \widehat{h}) = \inf_{\theta_i \in \Theta} Q_T^i(\theta_i, \widehat{h}) + o_p(1)$. The uniform consistency condition for the preliminary estimator (C3) was established in Lemma 3.5.1. Assumption (C4) is an identification condition that requires $\theta_{0i}$ to uniquely minimize $Q_0^i(\theta_i, h_0)$ over $\theta_i \in \Theta$. A necessary condition for identification (in a neighborhood of $\theta_{i0}$) is that the second derivative of the objective function with respect to $\theta_i$ has full rank. Consistency of the infeasible estimator $\widetilde{\theta}_i$ (assumption (C5)) that counterfactually assumes that the unobserved factors are known follows from standard arguments for extremum estimators (e.g. Wald (1949), Newey and McFadden (1994)).

**Theorem 3.5.1.** *Suppose that assumptions (C1)-(C6) hold. Then, as* $(T, N) \xrightarrow{j} \infty$, $\widehat{\theta}_i \xrightarrow{p} \theta_{0i}$ *for each fixed* $i$.

To derive the asymptotic distribution of the individual-specific estimators $\widehat{\theta}_i$, I assume that $\widehat{\theta}_i$ is consistent and that $\theta_{i0} \in \Theta$. In addition, I assume that:

(D1) $\frac{\partial Q_0}{\partial \theta}(\theta_{i0}, h_0) = 0$

(D2) For some sequence $\delta_T = o(1)$

$$\sup_{\|h - h_0\| \leq \delta_T} \sup_{\|\theta - \theta_{i0}\| \leq \delta_T} \left\| \frac{\partial^2 Q_T^i(\theta, h)}{\partial \theta \partial \theta^\intercal} - \frac{\partial^2 Q_T^i(\theta, h_0)}{\partial \theta \partial \theta^\intercal} \right\| = o_p(1)$$

where the $K_d + 2K_x \times K_d + 2K_x$ matrix $\frac{\partial^2 Q_T^i(\theta_0, h_0)}{\partial \theta \partial \theta^\intercal}$ has full rank.

(D3) There is a matrix $H_{i2}(\theta, h)$ of dimensions $K_d + 2K_x \times TK_x$ that satisfies for some sequence $\delta_T = o(1)$

$$\sup_{\|\theta_i - \theta_{i0}\| \leq \delta_T} \sup_{\|h - h_0\|_{\mathcal{H}} \leq \delta_T} \left\| \frac{\partial^2 Q_T^i}{\partial \theta \partial h^\intercal}(\theta_i, h) - H_{i2}(\theta_{i0}, h_0) \right\| \xrightarrow{p} 0,$$

where $\sup_{t \geq 1} ||H_{i2t}(\theta_{i0}, h_0)|| < C$ for each $K_d + 2K_x \times K_X$ submatrix $H_{i2t}(\theta_{i0}, h_0)$.

(D4) For some $J_i > 0$ and $J_i < C$

$$\sqrt{T} \frac{\partial Q_T^i}{\partial \theta}(\theta_{i0}, h_0) \xrightarrow{d} N(0, J_i)$$

(D5) $\frac{T(\log T)^2}{N} \to 0$

(D6) There is a $K_x \times K_x$ matrix

$$W_{jts}(\theta_{i0}, h) = \frac{\partial^3 Q_T^i(\theta_{i0}, h)}{\partial \theta_j \partial h_t^\intercal \partial h_s}$$

that satisfies for some sequence $\delta_T = o(1)$

$$\sup_{\|h-h_0\|\leq\delta_T} \|W_{jts}(\theta_{i0}, h)\| < C, \forall j, t, s$$

(D7) $\|\widehat{h} - h_0\|_{\mathcal{H}} = o_p(1)$

Assumption (D3) is a uniform convergence condition of the Hessian in a shrinking neighborhood of the true parameters $\theta_{i0}$ and $h_0$ and it can be replaced by a more primitive ULLN (Andrews (1993)). Assumption (D4) is analogous to asymptotic normality of the score and is satisfied because

$$\frac{\partial Q_T^i}{\partial\theta}(\theta_{i0}, h_0) = -\frac{1}{T}\sum_{t=1}^{T} \frac{Y_{it} - F_{it0}}{F_{it0}(1 - F_{it0})} f_{it0}[X_{it}^{\intercal} : d_t^{\intercal} : h_{0t}^{\intercal}]^{\intercal}$$

is a sample average with zero mean that is $IID$ conditional on the factors. The restriction on the relative size of $T$ and $N$ in (D5) ensures that the estimated preliminary functions $\widehat{h}$ do not affect the asymptotic distribution. Theorem 3.5.2 summarizes the asymptotic normality result for the individual-specific estimators $\widehat{\theta}_i$.

**Theorem 3.5.2.** *Suppose that assumptions (A2) and (D1)-(D7) hold. Then, as* $(T, N) \xrightarrow{j} \infty$,

$$\sqrt{T}(\widehat{\theta}_i - \theta_{0i}) \xrightarrow{d} N(0, V_i)$$

*for each fixed* $i$*, where:*

$$V_i = H_{1i}(\theta_{0i}, h_0)^{-1} J_i H_{1i}(\theta_{0i}, h_0)^{-1^{\intercal}}$$

$$H_{1i}(\theta_{i0}, h_0) = p\lim_{T\to\infty} \frac{\partial^2 Q_T^i(\theta_{i0}, h_0)}{\partial\theta\partial\theta^{\intercal}}.$$

### 3.5.2 Asymptotics for the mean group estimator

In this section, I investigate the asymptotic properties of the mean group estimator $\widehat{\beta}$ defined in equation (3.5) which is a subset of the parameter estimates contained in $\widehat{\theta}$. Consistency of $\widehat{\theta}$ follows by similar arguments as in the case of the individual-specific estimators $\widehat{\theta}_i$ and is summarized in the following theorem:

**Theorem 3.5.3.** *Suppose that assumptions (C1)-(C6) hold. Then, as* $(T, N) \xrightarrow{j} \infty$, $\widehat{\theta} \xrightarrow{p} \theta_0$.

To show that the mean group estimator is asymptotically normal, I assume that $\widehat{\theta}$ is consistent and asymptotically normal and that $\theta_0$ is an interior point in $\Theta$. Additionally, I assume that:

(E1) $\frac{1}{N}\sum\limits_{i=1}^{N} \frac{\partial Q_0}{\partial\theta}(\theta_{i0}, h_0) = 0$

(E2) There is a matrix $H_2(\theta, h)$ that satisfies for some $\delta_N = o(1)$

$$\sup_{\|\theta-\theta_0\|\leq\delta_N} \sup_{\|h-h_0\|_{\mathcal{H}}\leq\delta_N} \left\| \frac{1}{N}\sum_{i=1}^{N} \frac{\partial^2 \widehat{Q}_T^i}{\partial\theta\partial h^\intercal}(\theta, h) - H_2(\theta_0, h_0) \right\| \xrightarrow{p} 0$$

Assumptions E are similar to Assumptions D that have been imposed to establish asymptotic normality of the individual-specific estimators.

Theorem 3.5.4 contains the asymptotic normality result for the mean group estimator.

**Theorem 3.5.4.** *Suppose that assumptions (E1)-(E5), (D4) hold. Then, as* $(T, N) \xrightarrow{j} \infty$,

$$\sqrt{N}(\widehat{\beta} - \beta_0) \xrightarrow{d} N(0, \Sigma_\eta).$$

Observe that the asymptotic variance of the mean group estimator is equal to that of the random coefficients (assumption (A1)). In practice, $\Sigma_\eta$ can be consistently estimated by

$$\widehat{\Sigma}_\eta = \frac{1}{N-1}\sum_{i=1}^{N}(\widehat{\beta}_i - \widehat{\beta})(\widehat{\beta}_i - \widehat{\beta})^\intercal. \tag{3.14}$$

The estimator $\widehat{\Sigma}_\eta$ is identical to the one that is obtained in OLS and quantile regression settings (Pesaran (2006), Boneva et al. (2015)).

## 3.6 Small sample experiments

To complement the asymptotic analysis, this section studies the small sample properties of the *CCE mean group estimator* and compares them to the following set of alternative estimators:

1. The *infeasible mean group estimator* that counterfactually assumes that the unknown factors can be observed.

2. The *CCE mean group estimator with* $\overline{W}_t$ that approximates the unknown factors with the cross-sectional averages of the regressors $\overline{X}_t$ and instruments $\overline{W}_t$.

3. The *naive mean group estimator* that does not account for unobserved common factors.

4. The *linear probability mean group estimator* that replaces the probit model by a linear probability model.

The small sample performance of these estimators is evaluated in five experiments that cover a wide range of factor structures than can be encountered in economic and financial panel data sets: [18]

**Experiment 1** The data generating process (DGP) is

$$
\begin{aligned}
Y_{it}^* &= \alpha_i + \beta_{1i}X_{1it} + \beta_{2i}X_{2it} + \kappa_{1i}f_{1t} + \kappa_{2i}f_{2t} + \epsilon_{it}, \quad Y_{it} = I(Y_{it}^*) \\
X_{jit} &= a_{ji} + k_{ji1}f_{1t} + k_{ji2}f_{2t} + u_{jit}, \quad j = 1, 2 \\
W_{it} &= a_i^W + k_{i1}^W f_{1t} + k_{i2}^W f_{2t} + w_{it} \\
\epsilon_{it} &\sim NID(0, 1) \\
u_{jit} &\sim NID(0, 1), \quad j = 1, 2 \\
w_{it} &\sim NID(0, 1)
\end{aligned} \tag{3.15}
$$

where the factors are generated by[19]

$$
\begin{aligned}
f_{lt} =& \rho_f f_{lt-1} + \nu_{lft}, \quad t = -50, \ldots \\
&, T, l = 1, 2
\end{aligned} \tag{3.16}
$$

$$
\nu_{lft} \sim NID(\mu_f(1-\rho_f), 1-\rho_f^2), \quad \rho_f = 0.5, \mu_f = 0.5, l = 1, 2. \tag{3.17}
$$

The coefficients $\alpha_i$, $a_{ji}$ and $a_i^W$ are held fixed across replications and are initially generated as

$$
\begin{aligned}
\alpha_i &\sim NID(-0.5, 0.1) \\
a_{ji} &\sim NID(0.5, 0.1), \quad j = 1, 2 \\
a_i^W &\sim NID(0.5, 0.1).
\end{aligned} \tag{3.18}
$$

The remaining coefficients are drawn independently across replications according to

$$
\begin{aligned}
\beta_{1i} &= 0.5 + \eta_{1i}, \quad \eta_{1i} \sim NID(0, 0.02) \\
\beta_{2i} &= -0.5 + \eta_{2i}, \quad \eta_{2i} \sim NID(0, 0.02) \\
\kappa_{ij} &\sim NID(0.5, 0.1), \quad j = 1, 2 \\
k_{ji1} &\sim NID(0.5, 0.1), \quad j = 1, 2
\end{aligned}
$$

[18] When estimating binary choice models, one occasionally encounters the problem of quasi-complete separation. Quasi-complete separation occurs when the dependent variable separates the independent variables to certain degree. In that case, the maximum likelihood estimator does not exist and attempting to compute it usually results in an upward biased estimate. To mitigate this problem in the Monte Carlo experiments, I use the bias-reduction method of Firth (1993). Asymptotically, this estimator is equivalent to maximum likelihood to first order.

[19] The DGP for the factors (3.16) does not satisfy assumption (A2) because the factors are not bounded. But this does not affect the asymptotic theory because under normality as assumed here, the penalty term in the uniform rate (Lemma 3.5.1) is $\sqrt{\log(T)}$ which is smaller than the current penalty of $\log(T)$.

$$k_{i1}^W \sim NID(0.5, 0.1). \tag{3.19}$$

**Experiment 2** is identical to experiment 1 except that $\beta_{1i} = 0.5, \beta_{2i} = -0.5 \forall i$. There is no slope heterogeneity.

**Experiment 3** is identical to experiment 1 except that $k_{ji2} \sim NID(0, 0.1)$. The rank condition (3.8) is not satisfied.

**Experiment 4** is identical to experiment 1 except that

$$Y_{it}^* = \alpha_i + \beta_i X_{it} + \kappa_{1i} f_{1t} + \kappa_{2i} f_{2t} + \kappa_{3i} f_{3t} + \epsilon_{it}, \quad Y_{it} = I(Y_{it}^*) \tag{3.20}$$

where $\kappa_{i3}$ and $f_{t3}$ are generated as $\kappa_{i1}$ and $f_{t1}$. In this experiment, there are more unknown factors than proxies which illustrates another failure of the rank condition (3.8).

### 3.6.1 Coefficient estimates

To assess the small sample performance of the different estimators, I compute the maximal bias and RMSE for $\beta_1$ that are defined as:[20]

$$\text{RMSE}_\beta = \left( \sqrt{\frac{1}{R} \sum_{r=1}^{R} (\widehat{\beta}_{1r} - \beta_1)^2} \right)$$
$$\text{Bias}_\beta = \left( \beta_1 - \frac{1}{R} \sum_{r=1}^{R} \widehat{\beta}_{1r} \right),$$

where $R$ is the number of replications.

Tables 3.1-3.4 report RMSE and bias for experiments 1-4.[21] The naive estimator has poor small sample properties in all experimental settings. This result is not surprising because this estimator omits the unobserved common factors that play an important role in the DGP. In contrast, the CCE mean group estimator is comparable to the infeasible estimator in terms of RMSE even if the coefficients are homogeneous (Table 3.2). If the rank condition is not satisfied, the performance of the CCEMG estimator deteriorates (Tables 3.3 and 3.4).

The estimator that uses additional instruments $\overline{W}_t$ to approximate the unobserved factors performs similar to the CCE mean group estimator in terms of bias and RMSE with exception of experiment 3 where the estimator with $\overline{W}_t$ performs well despite of a failure of the rank condition (Table 3.3). In this situation, an instrument is available that is strongly correlated with the dependent variable. In

[20] Results for $\beta_2$ are almost identical.

[21] The linear probability estimator is excluded in this section because the coefficients represent marginal effects and are thus not comparable to the other estimates.

contrast, the estimator with $\overline{W}_t$ is biased in experiment 4. This experiment mimics a situation where the instruments are weak: in contrast to the other experiments, the instruments $W_{it}$ are generated by less factors than the dependent variable $Y_{it}$ which reduces the correlation between $\overline{W}_t$ and $\overline{Y}_t$.

Tables 3.1-3.4 also report empirical sizes and power. Power is computed under the alternative $\beta_1 = 0.45$ and the variance of $\widehat{\beta}_1$ is calculated using the formula in equation (3.14). While the naive estimator has distorted empirical sizes across all experiments, the empirical sizes of the CCE mean group estimator are close to the nominal size of 5% in all experiments except if the rank condition fails (Tables 3.3 and 3.4). With exception very small sample sizes, the CCE mean group estimator also has good power. The empirical sizes and powers of CCE mean group estimator with $\overline{W}_t$ are similar to those of the CCE mean group estimator with exception of experiment 3 where it outperforms the CCE mean group estimator (Table 3.3).

### 3.6.2 Marginal effects

Applied research usually reports marginal effects rather than coefficient estimates when estimating discrete choice models. Unlike coefficient estimates, marginal effects can be used to assess the economic significance of the results which is important to inform debates about economic policy. For the probit model, the average marginal effect is defined as:

$$ME_i = \beta_i \frac{1}{T} \sum_{t=1}^{T} \phi(\tilde{\alpha}_i^\intercal d_t + \beta_{1i} X_{1it} + \beta_{2i} X_{2it} + \tilde{\kappa}_i^\intercal \overline{X}_t).$$

Bias and RMSE for the marginal effect are computed as for the coefficient estimates in Section 3.6.1.

Tables 3.5-3.8 report RMSE and bias for marginal effects. Marginal effects computed from either CCE mean group estimates or CCE mean group estimates with $\overline{W}_t$ have similar bias and RMSE when compared to the infeasible marginal effects and outperform naive marginal effects that do not account for unobserved common factors. These conclusions hold even if the rank condition is not satisfied (Tables 3.7, 3.8).[22] The linear probability model augmented with cross-sectional averages has good small sample properties, too.[23]

Overall, the Monte Carlo evidence indicates that the CCE mean group estimator has good small sample properties compared to the infeasible estimator. These conclusions are robust to the case where coefficients are homogeneous.

[22] The robustness of marginal effects even if coefficient estimates are biased has been documented before, see e.g. Fernandez-Val and Weidner (2015).

[23] The linear probability model performs worse if the marginal effects at the average are computed instead of the average marginal effects. These results are available from the author on request.

## 3.7 The effect of corporate bond yields on bond issuance by US corporates[24]

At least since Modigliani and Miller (1958), the capital structure of firms has attracted much attention and there is a large empirical and theoretical literature that explores why the capital structure matters (Brealey et al. (2008)). For example, the mix of debt and equity is relevant in the presence of the bankruptcy costs or asymmetric information (Frank and Goyal (2008)).

Relative to equity, debt financing is an important source of external funds for US corporations (Denis and Mihov (2003)). Debt financing can take the form of bank loans, other loans or public debt. The focus here is on public debt. But in contrast to earlier studies (Frank and Goyal (2008)), I adopt an incremental approach that investigates the conditional probability of issuing a corporate bond which is particularly suitable for questions related to time-variation in the regressors. In my study, the yield paid by issuers in the corporate bond market is the regressor of primary interest.

Answering the question of how funding costs in corporate bond markets affect issuance decisions sheds light on a particular transmission mechanism of monetary policy: by means of conventional and unconventional monetary policy tools, the central bank can affect the interest rates firms face in corporate bond markets. Bond issuance, on the other hand, is often related to corporate investment and thus aggregate demand (Farrant et al. (2013)).

There is already a large literature that explores the determinants of bond issuance (e.g. Mizen and Tsoukas (2013), Badoer and James (2015), Adrian et al. (2012), Denis and Mihov (2003), Becker and Ivashina (2014)). These studies have documented that issuer characteristics like size, rating, profitability, leverage, equity prices, monetary policy and the supply of bank credit are important determinants of bond issuance. Other papers have investigated the effects of Quantitative Easing (Lo Duca et al. (2016)) or the Basel reforms on issuance decisions of banks or non-financial corporations (Baba and Inada (2009)). However, there is not much evidence yet on the effect of yields on bond issuances which is the contribution of this study. Additionally, previous studies have not controlled for common unobserved factors that affect both bond issuance and its determinants.

### 3.7.1 Data

The data set includes bond issuances by US corporates between 1990 and 2015 on a monthly frequency. The sample is restricted to bonds in US dollar, with a fixed coupon and short-run unsecured collateral. Non-bullet and callable bonds are

[24] I would like to thank Lu Liu, Menno Middeldorp and Magda Rutkowska for useful discussions about the empirical application and their help with the data.

excluded. The number of issuances is 5610 with an average size of approximately 300 million USD made by 1004 different firms. Time series of individual bond yields are obtained from Datastream and aggregated by issuer. Issuer-specific yields are constructed as the median of the individual bond yields. [25]

Figure 3.1 reports time series of the number of bond issuances and the average issuer-specific yield between 1990 and 2015. Over the sample period, the number of bond issuances increased and remained at high levels since 2003 with exception of a drop at the beginning of the financial crisis when yields increased sharply. The time series of yields and the number of issuances for financial sector firms co-move closely with the aggregate series. Albeit only one quarter of all firms are in the financial sector, a large number of issuances can be attributed to them. Figure 3.2 reports the cross-sectional mean, median and dispersion of yields over time. Yields exhibit a downward trend over the sample period. In 2008, both the level and the dispersion of yields increased sharply but started to fall again in 2009 which is in part explained by the Quantitative Easing program of the Federal Reserve.

Figure 3.3 illustrates the unconditional correlation between the cross-sectional average of yields and the number of issuances per month. For the pre-crisis period, there is a negative correlation for yields below 8. However, this correlation could be driven by common, unobserved shocks which will be controlled for in the regression analysis below. Finally, Figure 3.4 documents the number of issuances by firm. The distribution of the number of issuances is highly skewed with many firms only issuing one bond over the sample period: the average number of issuances is 6 but the median number of issuances is only 2.

### 3.7.2 Results

To investigate the effect of yields on bond issuance by US corporates, I estimate the econometric model in (3.1)-(3.3) where $Y_{it}$ indicates whether firm $i$ has issued a bond in time $t$ and $X_{it}$ contains the issuer's corporate bond yield and assets at the end of the previous month. The observed common factors $d_t$ include a constant, a measure of monetary policy and broker-dealer leverage which is a measure of bank credit conditions (Adrian et al. (2012)). For the pre-crisis period, the stance of monetary policy is measures by the federal funds rate, and in the post-crisis period, the change in Federal Reserve Holdings of Treasury Notes is used. In this specific empirical application, the unobserved factors can represent financial innovation that makes it easier for firms to tap the corporate bond market or policies that aim at deepening these markets, for example.

For the empirical analysis, the data set is restricted to firms with at least 30 time series observations and results are reported separately for the pre-and post

[25] This method ignores differences in duration and maturity across bond issuances. Constructing a better measure of issuer-specific yields is subject of ongoing work.

crisis period. In the post-crisis period, policies such as Quantitative Easing or credit guarantee schemes are likely to fundamentally change the incentives for corporates to issue bonds relative to the pre-crisis period.

Columns 1 to 3 in Table 3.9 report the mean group estimate of $\beta$ and marginal effects for the pre-crisis period. I find that the conditional probability of issuing a bond is higher if yields are low. This effect is statistically significant with exception of financial corporations but the marginal effects reveal that it is small in absolute magnitude. The effect of firm size is not statistically significant. For comparison, column 4 reports the mean group estimates when the common factors are omitted which differ from the CCE mean group estimates in size and statistical significance. Table 3.10 divides the pre-crisis sample in corporates with a low and high credit rating. With exception of financial firms where sample sizes are very small, yields are negatively related to the probability of issuing a bond for low-rated firms. In contrast, that effect is statistically insignificant conditional on a high credit rating.

In the post-crisis period, qualitatively the same observations can be made: for non-financial corporations, higher yields are associated with a less issuance activity (Tables 3.11). This result is driven by firms with a low credit rating (Table 3.12). Additionally, non-financial corporations that are relatively small are more likely to issue a bond. One explanation for this finding builds on the substitution from bank loans to bonds in the post-crisis period (Farrant et al. (2013)). This effect is likely to be stronger for relatively small firms that relied more heavily on bank loans prior to the financial crisis.

## 3.8 Conclusions

Economic variables are affected by common shocks such as financial crises, natural disasters, technological innovation or changes in the political or regulatory environment. These shocks tend to be difficult to measure and their impact differs across individual observations. As documented by Andrews (2005), common shocks create problems for inference if data are available for a single cross-sectional unit and the model is estimated by least squares or instrumental variable methods. But the increased availability of panel data where both the time series and cross-sectional dimensions are large offer new opportunities to control for these unobserved shocks (Bai (2009), Pesaran (2006)).

This chapter contributes to a growing literature on panel data models with cross-sectional dependence. The specific setting studied in this chapter is one where outcomes are discrete which introduces a nonlinearity. Discrete choice models are probably the most popular nonlinear panel data models in econometrics but the methodology developed here is applicable to nonlinear panel data models in general. The estimator I propose controls for unobserved common factors by means of cross-sectional averages of the regressors. The proposed estimator can be computed by estimating binary response models applied to regression that is augmented with the cross-sectional averages of the individual-specific regressors. The asymptotic properties of the individual-specific coefficients and their mean are documented. A Monte Carlo study assesses the behavior of the proposed methodology in small samples. The estimator is applied to the question of how of funding costs in corporate bond markets affect the decision to issue a corporate bond. I find that conditional probability of issuing a bond is larger in low yield environments for non-financial firms.

There are many ways in which this work can be developed further. An interesting extension of the empirical application is to examine how participation in a credit guarantee scheme affects the issuance decisions of corporates. These schemes were adopted in 2008 as part of financial sector rescue packages in order to help banks to retain access to funding markets (Grande et al. (2011)). In addition, I expect that constructing a firm-specific measure of credit supply from individual loan data can reveal additional insights on the substitution between bonds and loans. On the theoretical side, one area of future research is to extend the methodology proposed here to endogenous regressors or to homogeneous panels.

**Table 3.1:** Small sample properties of the mean group estimator $\widehat{\beta}$: Experiment 1

| T/N | Bias (× 1000) | | | | RMSE (× 1000) | | | | Power | | | | Size | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 50 | 100 | 200 | 300 | 50 | 100 | 200 | 300 | 50 | 100 | 200 | 300 | 50 | 100 | 200 | 300 |
| **Infeasible estimator** | | | | | | | | | | | | | | | | |
| 50 | 3.947 | 2.831 | 1.902 | 3.25 | 35.39 | 25.09 | 17.27 | 14.7 | 0.276 | 0.558 | 0.826 | 0.96 | 0.045 | 0.043 | 0.037 | 0.04 |
| 100 | -0.3039 | 0.4454 | 1.152 | 0.3768 | 22.8 | 16.59 | 11.3 | 9.548 | 0.522 | 0.848 | 0.995 | 1 | 0.047 | 0.047 | 0.039 | 0.044 |
| 200 | 0.1387 | 0.1602 | 0.3321 | 0.4307 | 15.71 | 11.1 | 7.975 | 6.496 | 0.868 | 0.993 | 1 | 1 | 0.04 | 0.044 | 0.04 | 0.046 |
| 300 | -0.5534 | 0.158 | 0.2649 | -0.1933 | 12.89 | 9.285 | 6.526 | 5.369 | 0.963 | 1 | 1 | 1 | 0.046 | 0.035 | 0.038 | 0.037 |
| **CCEMG estimator** | | | | | | | | | | | | | | | | |
| 50 | -0.7777 | -0.698 | -0.7033 | 0.5362 | 35.26 | 24.48 | 17.03 | 14.23 | 0.236 | 0.508 | 0.809 | 0.948 | 0.051 | 0.045 | 0.044 | 0.042 |
| 100 | -4.91 | -2.995 | -1.86 | -2.35 | 22.95 | 16.53 | 11.35 | 9.881 | 0.436 | 0.804 | 0.989 | 1 | 0.045 | 0.057 | 0.046 | 0.054 |
| 200 | -4.564 | -3.678 | -2.515 | -2.37 | 16.26 | 11.72 | 8.398 | 6.868 | 0.795 | 0.987 | 1 | 1 | 0.046 | 0.055 | 0.065 | 0.056 |
| 300 | -5.24 | -3.36 | -2.634 | -2.89 | 13.72 | 9.865 | 7.055 | 6.024 | 0.923 | 0.998 | 1 | 1 | 0.057 | 0.058 | 0.061 | 0.073 |
| **CCEMG estimator with $\overline{W}$** | | | | | | | | | | | | | | | | |
| 50 | 0.599 | 0.02083 | -0.2738 | 1.314 | 36.18 | 24.84 | 17.13 | 14.59 | 0.234 | 0.498 | 0.793 | 0.949 | 0.052 | 0.053 | 0.046 | 0.042 |
| 100 | -4.078 | -2.469 | -1.457 | -2.043 | 23.03 | 16.76 | 11.44 | 9.887 | 0.446 | 0.804 | 0.99 | 1 | 0.047 | 0.058 | 0.045 | 0.055 |
| 200 | -3.679 | -3.251 | -2.26 | -2.187 | 16.16 | 11.66 | 8.34 | 6.832 | 0.806 | 0.985 | 1 | 1 | 0.045 | 0.052 | 0.063 | 0.059 |
| 300 | -4.413 | -2.953 | -2.386 | -2.743 | 13.47 | 9.786 | 6.974 | 5.967 | 0.933 | 0.998 | 1 | 1 | 0.056 | 0.051 | 0.054 | 0.067 |
| **Naive estimator** | | | | | | | | | | | | | | | | |
| 50 | 159.4 | 158.7 | 158.7 | 158.2 | 163.4 | 161.3 | 160.5 | 159.7 | 1 | 1 | 1 | 1 | 0.998 | 1 | 1 | 1 |
| 100 | 157.4 | 157.8 | 157.9 | 157.8 | 159.1 | 158.9 | 158.7 | 158.5 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 200 | 158.4 | 157.7 | 158.2 | 158.6 | 159.2 | 158.3 | 158.6 | 158.9 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 300 | 158.3 | 158.4 | 158.3 | 157.7 | 158.9 | 158.8 | 158.5 | 157.9 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

*Notes:* The mean group estimator is defined in (3.5) and the data generating process in (3.15)-(3.19). The nominal size is 5% and power is computed under the alternative $\beta_1 = 0.45$. The number of replications is set to 1000.

**Table 3.2:** Small sample properties of the mean group estimator $\widehat{\beta}$: Experiment 2

| T/N | Bias (× 1000) | | | | RMSE (× 1000) | | | | Power | | | | Size | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 50 | 100 | 200 | 300 | 50 | 100 | 200 | 300 | 50 | 100 | 200 | 300 | 50 | 100 | 200 | 300 |
| **Infeasible estimator** | | | | | | | | | | | | | | | | |
| 50 | 2.848 | 1.518 | 2.094 | 2.456 | 35.26 | 24.73 | 17.3 | 14.81 | 0.28 | 0.511 | 0.839 | 0.951 | 0.046 | 0.038 | 0.043 | 0.05 |
| 100 | -0.9409 | 0.2678 | 0.5554 | 0.473 | 23.04 | 16.1 | 11.96 | 9.294 | 0.512 | 0.861 | 0.99 | 1 | 0.047 | 0.035 | 0.058 | 0.048 |
| 200 | -0.2583 | -0.1573 | 0.04374 | 0.1231 | 16.15 | 11.29 | 8.073 | 6.608 | 0.853 | 0.994 | 1 | 1 | 0.042 | 0.047 | 0.05 | 0.045 |
| 300 | 0.04022 | 0.3875 | 0.1868 | -0.05547 | 12.28 | 8.838 | 6.341 | 5.108 | 0.98 | 1 | 1 | 1 | 0.037 | 0.04 | 0.041 | 0.037 |
| **CCEMG estimator** | | | | | | | | | | | | | | | | |
| 50 | -1.784 | -1.505 | -1.041 | -0.02763 | 34.44 | 24.2 | 16.99 | 14.34 | 0.236 | 0.463 | 0.805 | 0.939 | 0.05 | 0.055 | 0.04 | 0.052 |
| 100 | -5.404 | -3.303 | -2.29 | -2.163 | 23.72 | 16.22 | 11.91 | 9.452 | 0.448 | 0.812 | 0.985 | 1 | 0.056 | 0.049 | 0.055 | 0.051 |
| 200 | -4.977 | -3.716 | -2.91 | -2.599 | 16.91 | 11.89 | 8.569 | 7.101 | 0.787 | 0.989 | 1 | 1 | 0.064 | 0.056 | 0.073 | 0.079 |
| 300 | -4.805 | -3.101 | -2.823 | -2.777 | 13.45 | 9.45 | 6.907 | 5.861 | 0.942 | 0.999 | 1 | 1 | 0.062 | 0.059 | 0.066 | 0.079 |
| **CCEMG estimator with $\overline{W}$** | | | | | | | | | | | | | | | | |
| 50 | -0.6561 | -0.7614 | -0.3792 | 0.6817 | 35.28 | 24.48 | 17.34 | 14.76 | 0.232 | 0.459 | 0.798 | 0.938 | 0.055 | 0.044 | 0.038 | 0.046 |
| 100 | -4.396 | -2.732 | -1.939 | -1.872 | 23.81 | 16.32 | 11.92 | 9.431 | 0.471 | 0.817 | 0.985 | 1 | 0.058 | 0.046 | 0.054 | 0.047 |
| 200 | -4.075 | -3.261 | -2.658 | -2.404 | 16.86 | 11.83 | 8.49 | 7.057 | 0.799 | 0.989 | 1 | 1 | 0.065 | 0.058 | 0.069 | 0.076 |
| 300 | -3.99 | -2.685 | -2.557 | -2.597 | 13.2 | 9.369 | 6.823 | 5.782 | 0.952 | 0.999 | 1 | 1 | 0.055 | 0.052 | 0.067 | 0.079 |
| **Naive estimator** | | | | | | | | | | | | | | | | |
| 50 | 158.4 | 157.6 | 156.2 | 158 | 161.9 | 160.2 | 157.9 | 159.5 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 100 | 156.1 | 157.5 | 157.3 | 157.5 | 157.9 | 158.6 | 158.1 | 158.2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 200 | 157.9 | 157.6 | 158.2 | 158.1 | 158.8 | 158.1 | 158.5 | 158.4 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 300 | 158.2 | 158.4 | 158.3 | 158.2 | 158.7 | 158.7 | 158.5 | 158.4 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

*Notes:* The mean group estimator is defined in (3.5) and the data generating process in (3.15)-(3.19) except that $\beta_{1i} = 0.5, \forall i$. The nominal size is 5% and power is computed under the alternative $\beta_1 = 0.45$. The number of replications is set to 1000.

**Table 3.3:** Small sample properties of the mean group estimator $\widehat{\beta}$: Experiment 3

| T/N | Bias (× 1000) | | | | RMSE (× 1000) | | | | Power | | | | Size | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 50 | 100 | 200 | 300 | 50 | 100 | 200 | 300 | 50 | 100 | 200 | 300 | 50 | 100 | 200 | 300 |
| | | | | | | | | Infeasible estimator | | | | | | | | |
| 50 | 2.456 | 2.582 | 2.393 | 2.867 | 35.47 | 25.88 | 17.79 | 14.51 | 0.267 | 0.547 | 0.839 | 0.963 | 0.051 | 0.053 | 0.046 | 0.042 |
| 100 | 0.8102 | 0.6675 | 1.358 | 0.7839 | 23.8 | 15.94 | 11.56 | 9.35 | 0.534 | 0.868 | 0.995 | 1 | 0.047 | 0.028 | 0.047 | 0.042 |
| 200 | 1.054 | 0.3152 | -0.1941 | 0.03611 | 16.28 | 11.74 | 8 | 6.61 | 0.865 | 0.992 | 1 | 1 | 0.043 | 0.052 | 0.048 | 0.035 |
| 300 | 0.1421 | 0.4703 | 0.06673 | 0.13 | 12.72 | 9.217 | 6.451 | 5.344 | 0.974 | 1 | 1 | 1 | 0.036 | 0.043 | 0.036 | 0.047 |
| | | | | | | | | CCEMG estimator | | | | | | | | |
| 50 | -47.42 | -46.41 | -46.59 | -45.1 | 57.65 | 52.3 | 49.65 | 47.59 | 0.041 | 0.046 | 0.055 | 0.081 | 0.257 | 0.481 | 0.762 | 0.856 |
| 100 | -49.9 | -49.36 | -48.45 | -48.62 | 54.68 | 51.74 | 49.96 | 49.67 | 0.026 | 0.026 | 0.051 | 0.052 | 0.501 | 0.817 | 0.97 | 0.998 |
| 200 | -50.8 | -50.53 | -50.48 | -50.09 | 53.06 | 51.74 | 51.16 | 50.56 | 0.01 | 0.011 | 0.021 | 0.028 | 0.766 | 0.98 | 1 | 1 |
| 300 | -51.69 | -50.5 | -50.82 | -50.34 | 53.12 | 51.25 | 51.25 | 50.65 | 0.005 | 0.012 | 0.016 | 0.016 | 0.911 | 0.999 | 1 | 1 |
| | | | | | | | | CCEMG estimator with $\overline{W}$ | | | | | | | | |
| 50 | -1.72 | 0.5095 | 1.569 | 2.573 | 35.89 | 26.03 | 18.14 | 14.75 | 0.23 | 0.478 | 0.832 | 0.951 | 0.046 | 0.045 | 0.049 | 0.038 |
| 100 | -3.459 | -1.789 | 0.2207 | 0.03279 | 24.43 | 16 | 11.58 | 9.353 | 0.457 | 0.824 | 0.991 | 1 | 0.051 | 0.032 | 0.046 | 0.039 |
| 200 | -3.883 | -2.221 | -1.446 | -0.7808 | 16.87 | 12.01 | 8.176 | 6.68 | 0.772 | 0.983 | 1 | 1 | 0.047 | 0.055 | 0.051 | 0.037 |
| 300 | -4.742 | -2.005 | -1.232 | -0.7198 | 13.56 | 9.452 | 6.571 | 5.405 | 0.933 | 0.999 | 1 | 1 | 0.058 | 0.053 | 0.043 | 0.044 |
| | | | | | | | | Naive estimator | | | | | | | | |
| 50 | 56.14 | 57.36 | 56.35 | 58.45 | 66.39 | 63.84 | 61.45 | 63.34 | 0.907 | 0.991 | 0.999 | 1 | 0.432 | 0.673 | 0.851 | 0.919 |
| 100 | 56.09 | 55.97 | 55.6 | 56.08 | 60.91 | 59.06 | 58.14 | 58.4 | 0.996 | 1 | 1 | 1 | 0.651 | 0.899 | 0.977 | 0.99 |
| 200 | 56.72 | 55.15 | 55.2 | 55.38 | 59.23 | 56.84 | 56.35 | 56.41 | 1 | 1 | 1 | 1 | 0.886 | 0.986 | 0.999 | 1 |
| 300 | 55.82 | 55.86 | 55.19 | 54.89 | 57.47 | 56.84 | 55.97 | 55.61 | 1 | 1 | 1 | 1 | 0.963 | 1 | 1 | 1 |

*Notes:* The mean group estimator is defined in (3.5) and the data generating process in (3.15)-(3.19) except that $k_{ji2} \sim NID(0, 0.1)$. The nominal size is 5% and power is computed under the alternative $\beta_1 = 0.45$. The number of replications is set to 1000.

**Table 3.4:** Small sample properties of the mean group estimator $\widehat{\beta}$: Experiment 4

| T/N | Bias ($\times$ 1000) | | | | RMSE ($\times$ 1000) | | | | Power | | | | Size | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 50 | 100 | 200 | 300 | 50 | 100 | 200 | 300 | 50 | 100 | 200 | 300 | 50 | 100 | 200 | 300 |
| Infeasible estimator | | | | | | | | | | | | | | | | |
| 50 | 3.073 | 1.965 | 2.254 | 1.226 | 37.17 | 26.36 | 18.29 | 14.44 | 0.239 | 0.476 | 0.808 | 0.917 | 0.044 | 0.046 | 0.033 | 0.032 |
| 100 | 0.2937 | 1.201 | 0.4822 | 0.7025 | 24.5 | 16.99 | 12.69 | 9.811 | 0.515 | 0.828 | 0.985 | 0.999 | 0.05 | 0.041 | 0.059 | 0.041 |
| 200 | -0.1682 | -0.9886 | 0.6583 | 0.2589 | 16.72 | 11.87 | 8.387 | 7.068 | 0.82 | 0.988 | 1 | 1 | 0.04 | 0.041 | 0.036 | 0.053 |
| 300 | -0.1418 | 0.2268 | -0.2255 | -0.1925 | 13.21 | 9.568 | 6.746 | 5.646 | 0.961 | 0.999 | 1 | 1 | 0.042 | 0.039 | 0.044 | 0.05 |
| 300 | -0.5534 | 0.158 | 0.2649 | -0.1933 | 12.89 | 9.285 | 6.526 | 5.369 | 0.963 | 1 | 1 | 1 | 0.046 | 0.035 | 0.038 | 0.037 |
| CCEMG estimator | | | | | | | | | | | | | | | | |
| 50 | -50.76 | -51.18 | -49.94 | -51.33 | 60.48 | 57.1 | 53.32 | 53.82 | 0.046 | 0.062 | 0.066 | 0.095 | 0.329 | 0.567 | 0.788 | 0.923 |
| 100 | -55.86 | -54.05 | -54.37 | -53.56 | 60.29 | 56.38 | 55.94 | 54.67 | 0.063 | 0.056 | 0.11 | 0.114 | 0.66 | 0.905 | 0.992 | 0.999 |
| 200 | -56.63 | -56.4 | -54.43 | -54.81 | 58.77 | 57.54 | 55.11 | 55.36 | 0.067 | 0.095 | 0.103 | 0.158 | 0.935 | 0.998 | 1 | 1 |
| 300 | -56.67 | -55.36 | -55.3 | -55.35 | 58.03 | 56.12 | 55.73 | 55.68 | 0.069 | 0.089 | 0.149 | 0.198 | 0.989 | 1 | 1 | 1 |
| CCEMG estimator with $\overline{W}$ | | | | | | | | | | | | | | | | |
| 50 | -49.67 | -49.86 | -49.2 | -50.75 | 60.02 | 56.12 | 52.71 | 53.32 | 0.04 | 0.063 | 0.065 | 0.092 | 0.309 | 0.539 | 0.769 | 0.915 |
| 100 | -55.07 | -53.51 | -53.98 | -53.29 | 59.57 | 55.87 | 55.59 | 54.41 | 0.056 | 0.056 | 0.109 | 0.118 | 0.643 | 0.897 | 0.989 | 0.998 |
| 200 | -55.99 | -56.09 | -54.24 | -54.66 | 58.19 | 57.24 | 54.93 | 55.21 | 0.064 | 0.091 | 0.103 | 0.161 | 0.928 | 0.998 | 1 | 1 |
| 300 | -56.08 | -55.04 | -55.1 | -55.22 | 57.44 | 55.81 | 55.53 | 55.56 | 0.062 | 0.083 | 0.141 | 0.194 | 0.987 | 1 | 1 | 1 |
| Naive estimator | | | | | | | | | | | | | | | | |
| 50 | 103.9 | 103.4 | 100.8 | 102.1 | 110.4 | 108.2 | 104.9 | 105.7 | 0.998 | 1 | 1 | 1 | 0.889 | 0.976 | 0.993 | 0.997 |
| 100 | 99.93 | 101.6 | 100.5 | 100.7 | 103.4 | 103.8 | 102.5 | 102.3 | 1 | 1 | 1 | 1 | 0.993 | 0.999 | 1 | 1 |
| 200 | 99.69 | 99.92 | 100.9 | 101.2 | 101.4 | 101 | 101.9 | 102 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 300 | 100.8 | 100 | 100.9 | 100.2 | 101.9 | 100.8 | 101.4 | 100.8 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

*Notes:* The mean group estimator is defined in (3.5) and the data generating process in (3.16)-(3.19) and (3.20). The nominal size is 5% and power is computed under the alternative $\beta_1 = 0.45$. The number of replications is set to 1000.

**Table 3.5:** Small sample properties of the marginal effect $\widehat{ME}$: Experiment 1

| T/N | Bias (× 1000) | | | | RMSE (× 1000) | | | |
|---|---|---|---|---|---|---|---|---|
| | 50 | 100 | 200 | 300 | 50 | 100 | 200 | 300 |
| | Infeasible estimator | | | | | | | |
| 50 | -4.493 | -4.719 | -4.979 | -4.644 | 9.805 | 7.8 | 6.531 | 5.848 |
| 100 | -2.692 | -2.537 | -2.308 | -2.541 | 6.626 | 5.056 | 3.777 | 3.58 |
| 200 | -1.294 | -1.265 | -1.225 | -1.196 | 4.445 | 3.284 | 2.488 | 2.131 |
| 300 | -1.024 | -0.8424 | -0.7979 | -0.9292 | 3.682 | 2.678 | 1.959 | 1.733 |
| | CCEMG estimator | | | | | | | |
| 50 | -4.66 | -4.805 | -4.993 | -4.721 | 9.959 | 7.796 | 6.534 | 5.897 |
| 100 | -2.77 | -2.55 | -2.373 | -2.591 | 6.604 | 5.046 | 3.811 | 3.63 |
| 200 | -1.34 | -1.383 | -1.237 | -1.242 | 4.488 | 3.349 | 2.513 | 2.163 |
| 300 | -1.085 | -0.8714 | -0.8167 | -0.9492 | 3.687 | 2.713 | 1.988 | 1.73 |
| | CCEMG estimator with $\overline{W}$ | | | | | | | |
| 50 | -5.647 | -5.855 | -6.066 | -5.75 | 10.53 | 8.489 | 7.397 | 6.772 |
| 100 | -3.403 | -3.179 | -2.975 | -3.194 | 6.911 | 5.435 | 4.226 | 4.087 |
| 200 | -1.649 | -1.717 | -1.567 | -1.57 | 4.595 | 3.503 | 2.691 | 2.368 |
| 300 | -1.296 | -1.093 | -1.035 | -1.176 | 3.76 | 2.8 | 2.087 | 1.866 |
| | Naive estimator | | | | | | | |
| 50 | 58.72 | 58.69 | 58.72 | 58.45 | 59.54 | 59.25 | 59.14 | 58.83 |
| 100 | 62.32 | 62.48 | 62.46 | 62.46 | 62.68 | 62.71 | 62.66 | 62.63 |
| 200 | 64.45 | 64.31 | 64.44 | 64.58 | 64.62 | 64.44 | 64.53 | 64.65 |
| 300 | 65.08 | 65.12 | 65.08 | 64.91 | 65.19 | 65.2 | 65.14 | 64.97 |
| | Linear probability estimator | | | | | | | |
| 50 | 5.597 | 5.295 | 5.031 | 5.31 | 10.82 | 8.35 | 6.71 | 6.484 |
| 100 | 2.568 | 2.686 | 2.862 | 2.657 | 6.798 | 5.286 | 4.224 | 3.728 |
| 200 | 1.355 | 1.326 | 1.477 | 1.445 | 4.618 | 3.411 | 2.719 | 2.342 |
| 300 | 0.7102 | 0.9556 | 1.006 | 0.8585 | 3.719 | 2.807 | 2.126 | 1.72 |

*Notes:* The mean group estimator of the average marginal effect is reported. The data generating process in (3.15)-(3.19). The number of replications is set to 1000.

**Table 3.6:** Small sample properties of the marginal effect $\widehat{ME}$: Experiment 2

| T/N | Bias (× 1000) | | | | RMSE (× 1000) | | | |
|---|---|---|---|---|---|---|---|---|
| | 50 | 100 | 200 | 300 | 50 | 100 | 200 | 300 |
| | Infeasible estimator | | | | | | | |
| 50 | -4.692 | -5.052 | -4.835 | -4.808 | 9.854 | 7.935 | 6.483 | 6.022 |
| 100 | -2.917 | -2.557 | -2.488 | -2.529 | 6.794 | 4.964 | 4.047 | 3.541 |
| 200 | -1.41 | -1.343 | -1.287 | -1.287 | 4.612 | 3.36 | 2.551 | 2.208 |
| 300 | -0.8469 | -0.7705 | -0.8212 | -0.8939 | 3.466 | 2.548 | 1.914 | 1.661 |
| | CCEMG estimator | | | | | | | |
| 50 | -4.756 | -4.992 | -4.952 | -4.803 | 9.786 | 7.843 | 6.546 | 5.996 |
| 100 | -2.94 | -2.609 | -2.505 | -2.558 | 6.86 | 4.958 | 4.016 | 3.557 |
| 200 | -1.464 | -1.394 | -1.33 | -1.318 | 4.671 | 3.406 | 2.585 | 2.234 |
| 300 | -0.9196 | -0.7918 | -0.8693 | -0.9161 | 3.605 | 2.594 | 1.945 | 1.69 |
| | CCEMG estimator with $\overline{W}$ | | | | | | | |
| 50 | -5.804 | -6.007 | -6.014 | -5.832 | 10.43 | 8.557 | 7.397 | 6.875 |
| 100 | -3.53 | -3.219 | -3.128 | -3.175 | 7.163 | 5.325 | 4.431 | 4.027 |
| 200 | -1.775 | -1.726 | -1.66 | -1.648 | 4.794 | 3.564 | 2.765 | 2.445 |
| 300 | -1.139 | -1.015 | -1.088 | -1.138 | 3.661 | 2.681 | 2.052 | 1.819 |
| | Naive estimator | | | | | | | |
| 50 | 58.66 | 58.24 | 57.95 | 58.51 | 59.38 | 58.8 | 58.38 | 58.87 |
| 100 | 61.96 | 62.48 | 62.32 | 62.35 | 62.32 | 62.72 | 62.52 | 62.52 |
| 200 | 64.4 | 64.32 | 64.46 | 64.42 | 64.57 | 64.44 | 64.55 | 64.51 |
| 300 | 65.09 | 65.06 | 65.09 | 65.09 | 65.19 | 65.13 | 65.15 | 65.14 |
| | Linear probability estimator | | | | | | | |
| 50 | 5.395 | 4.987 | 5.23 | 5.221 | 10.61 | 8.089 | 6.97 | 6.444 |
| 100 | 2.389 | 2.64 | 2.753 | 2.738 | 6.861 | 5.093 | 4.262 | 3.731 |
| 200 | 1.26 | 1.294 | 1.362 | 1.371 | 4.721 | 3.451 | 2.68 | 2.306 |
| 300 | 0.9391 | 1.033 | 0.9373 | 0.8923 | 3.663 | 2.752 | 2.02 | 1.707 |

*Notes:* The mean group estimator of the average marginal effect is reported. The data generating process in (3.15)-(3.19) except that $\beta_{1i} = 0.5, \forall i$. The number of replications is set to 1000.

**Table 3.7:** Small sample properties of the marginal effect $\widehat{ME}$: Experiment 3

| T/N | Bias ($\times$ 1000) | | | | RMSE ($\times$ 1000) | | | |
|---|---|---|---|---|---|---|---|---|
| | 50 | 100 | 200 | 300 | 50 | 100 | 200 | 300 |
| | Infeasible estimator | | | | | | | |
| 50 | -4.772 | -4.809 | -4.849 | -4.719 | 9.979 | 7.911 | 6.549 | 5.891 |
| 100 | -2.432 | -2.463 | -2.286 | -2.444 | 6.774 | 4.916 | 3.816 | 3.47 |
| 200 | -1.018 | -1.204 | -1.364 | -1.299 | 4.533 | 3.422 | 2.572 | 2.226 |
| 300 | -0.8342 | -0.7421 | -0.856 | -0.8362 | 3.576 | 2.625 | 1.964 | 1.684 |
| | CCEMG estimator | | | | | | | |
| 50 | -6.241 | -6.193 | -6.335 | -6.111 | 10.72 | 8.757 | 7.691 | 7.074 |
| 100 | -3.477 | -3.55 | -3.389 | -3.475 | 7.345 | 5.563 | 4.61 | 4.261 |
| 200 | -1.953 | -2.055 | -2.228 | -2.156 | 4.964 | 3.782 | 3.152 | 2.823 |
| 300 | -1.623 | -1.513 | -1.654 | -1.602 | 3.938 | 3.005 | 2.457 | 2.183 |
| | CCEMG estimator with $\overline{W}$ | | | | | | | |
| 50 | -5.933 | -5.896 | -5.936 | -5.779 | 10.65 | 8.654 | 7.421 | 6.79 |
| 100 | -3.064 | -3.134 | -2.921 | -3.069 | 7.147 | 5.285 | 4.241 | 3.94 |
| 200 | -1.461 | -1.572 | -1.712 | -1.644 | 4.73 | 3.578 | 2.78 | 2.447 |
| 300 | -1.135 | -0.9892 | -1.104 | -1.073 | 3.682 | 2.718 | 2.084 | 1.816 |
| | Naive estimator | | | | | | | |
| 50 | 36.29 | 36.76 | 36.51 | 36.93 | 37.98 | 37.91 | 37.57 | 37.96 |
| 100 | 40.13 | 40.14 | 39.92 | 40.11 | 40.91 | 40.7 | 40.41 | 40.59 |
| 200 | 42.18 | 41.73 | 41.76 | 41.75 | 42.59 | 42.03 | 41.99 | 41.96 |
| 300 | 42.57 | 42.54 | 42.47 | 42.22 | 42.83 | 42.72 | 42.64 | 42.36 |
| | Linear probability estimator | | | | | | | |
| 50 | 4.417 | 4.362 | 4.256 | 4.429 | 10.22 | 7.913 | 6.314 | 5.825 |
| 100 | 2.111 | 2.089 | 2.243 | 2.133 | 6.985 | 4.957 | 3.929 | 3.342 |
| 200 | 0.95 | 0.8226 | 0.6537 | 0.7341 | 4.763 | 3.334 | 2.368 | 2.013 |
| 300 | 0.3361 | 0.4472 | 0.2769 | 0.3371 | 3.666 | 2.694 | 1.863 | 1.539 |

*Notes:* The mean group estimator of the average marginal effect is reported. The data generating process in (3.15)-(3.19) except that $k_{ji2} \sim NID(0, 0.1)$. The number of replications is set to 1000.

**Table 3.8:** Small sample properties of the marginal effect $\widehat{ME}$: Experiment 4

| T/N | Bias (× 1000) | | | | RMSE (× 1000) | | | |
|---|---|---|---|---|---|---|---|---|
| | 50 | 100 | 200 | 300 | 50 | 100 | 200 | 300 |
| | | | | Infeasible estimator | | | | |
| 50 | -4.743 | -4.935 | -4.886 | -5.094 | 9.767 | 7.852 | 6.417 | 6.088 |
| 100 | -2.571 | -2.389 | -2.552 | -2.476 | 6.539 | 4.8 | 4.027 | 3.445 |
| 200 | -1.33 | -1.549 | -1.118 | -1.235 | 4.432 | 3.378 | 2.388 | 2.154 |
| 300 | -0.8832 | -0.8189 | -0.9177 | -0.9145 | 3.476 | 2.571 | 1.944 | 1.705 |
| | | | | CCEMG estimator | | | | |
| 50 | -4.791 | -4.977 | -4.819 | -5.15 | 9.53 | 7.837 | 6.319 | 6.125 |
| 100 | -2.593 | -2.442 | -2.587 | -2.485 | 6.605 | 4.795 | 4.063 | 3.445 |
| 200 | -1.4 | -1.566 | -1.134 | -1.253 | 4.497 | 3.368 | 2.418 | 2.19 |
| 300 | -0.9236 | -0.8231 | -0.9073 | -0.9269 | 3.557 | 2.622 | 1.985 | 1.718 |
| | | | | CCEMG estimator with $\overline{W}$ | | | | |
| 50 | -5.82 | -5.889 | -5.839 | -6.156 | 10.17 | 8.501 | 7.129 | 6.993 |
| 100 | -3.197 | -3.024 | -3.189 | -3.071 | 6.841 | 5.121 | 4.489 | 3.888 |
| 200 | -1.727 | -1.89 | -1.461 | -1.577 | 4.622 | 3.536 | 2.586 | 2.395 |
| 300 | -1.136 | -1.038 | -1.126 | -1.146 | 3.615 | 2.699 | 2.092 | 1.845 |
| | | | | Naive estimator | | | | |
| 50 | 54.39 | 54.38 | 53.31 | 54.09 | 55.5 | 55.22 | 54.09 | 54.83 |
| 100 | 57.76 | 58.14 | 57.96 | 57.88 | 58.37 | 58.57 | 58.34 | 58.21 |
| 200 | 59.55 | 59.67 | 59.98 | 60.09 | 59.84 | 59.87 | 60.17 | 60.26 |
| 300 | 60.54 | 60.21 | 60.59 | 60.39 | 60.72 | 60.35 | 60.7 | 60.5 |
| | | | | Linear probability estimator | | | | |
| 50 | 4.826 | 4.529 | 4.796 | 4.322 | 9.904 | 7.927 | 6.499 | 5.603 |
| 100 | 2.434 | 2.58 | 2.453 | 2.539 | 6.65 | 4.97 | 4.075 | 3.555 |
| 200 | 1.242 | 1.027 | 1.457 | 1.316 | 4.592 | 3.198 | 2.648 | 2.253 |
| 300 | 0.8096 | 0.9192 | 0.8318 | 0.814 | 3.636 | 2.697 | 1.997 | 1.684 |

*Notes:* The mean group estimator of the average marginal effect is is reported. The data generating process in (3.16)-(3.19) and (3.20). The number of replications is set to 1000.

**Table 3.9:** The effect of yields on bond issuance for US corporates in the pre-crisis period

| | All | Financial | Other | All (no factors) |
|---|---|---|---|---|
| Coefficient estimates | | | | |
| Yield | -0.162 | -0.148 | -0.217 | -0.091 |
| | (-1.938) | (-0.853) | (-2.122) | (-2.653) |
| Size | 0.064 | 0.006 | 0.064 | 0.091 |
| | (0.281) | (0.168) | (0.195) | (0.543) |
| Marginal effects | | | | |
| Yield | -0.018 | -0.006 | -0.025 | -0.013 |
| Size | 0.015 | -0.003 | 0.02 | 0.021 |
| Observations | 321 | 62 | 221 | 321 |

*Notes:* The dependent variable is 1 if a firm issues a bond in a particular month and zero otherwise. Yield is the the firm-specific corporate bond yield and size is measured by assets. All specification include a measure of credit supply (leverage in the broker-dealer market) and the federal funds rate as a common factor. The first column uses all firms, the second column uses financial sector firms and the third column uses all other firms (excluding mining and agriculture). The last column reports the results when the common unobserved factors are omitted. t-statistics are shown in parenthesis.

**Table 3.10:** The effect of yields on bond issuance for US corporates in the pre-crisis period by credit rating

| | All | | Financial | | Other | |
|---|---|---|---|---|---|---|
| | High | Low | High | Low | High | Low |
| Coefficient estimates | | | | | | |
| Yield | 0.017 | -0.224 | -0.185 | -0.026 | 0.042 | -0.258 |
| | (0.177) | (-2.404) | (-0.79) | (-0.175) | (0.409) | (-1.929) |
| Size | 0.009 | -0.174 | 0.062 | 0.006 | 0.013 | -0.272 |
| | (0.091) | (-1.065) | (1.104) | (0.298) | (0.1) | (-1.065) |
| Marginal effects | | | | | | |
| Yield | 0.001 | -0.018 | -0.023 | 0.01 | 0.004 | -0.02 |
| Size | -0.005 | -0.024 | 0.006 | 0 | -0.006 | -0.037 |
| Observations | 135 | 135 | 27 | 28 | 99 | 85 |

*Notes:* The dependent variable is 1 if a firm issues a bond in a particular month and zero otherwise. Yield is the firm-specific corporate bond yield and size is measured by assets. All specification include a measure of credit supply (leverage in the broker-dealer market) and the federal funds rate as a common factor. Columns 1-2 use all firms, columns 3-4 use financial sector firms and columns 5-6 use all other firms (excluding mining and agriculture). Low (high) means that the issuer has a credit rating below (above) the sample median. t-statistics are shown in parenthesis.

**Table 3.11:** The effect of yields on bond issuance for US corporates in the post-crisis period

| | All | Financial | Other | All (no factors) |
|---|---|---|---|---|
| Coefficient estimates | | | | |
| Yield | -0.043 | -0.004 | -0.099 | 0.011 |
| | (-1.056) | (-0.074) | (-1.894) | (0.47) |
| Size | -0.194 | -0.028 | -0.273 | -0.115 |
| | (-2.62) | (-0.488) | (-2.683) | (-1.945) |
| Marginal effects | | | | |
| Yield | -0.007 | 0.003 | -0.015 | -0.002 |
| Size | -0.024 | -0.006 | -0.032 | -0.015 |
| Observations | 378 | 72 | 266 | 378 |

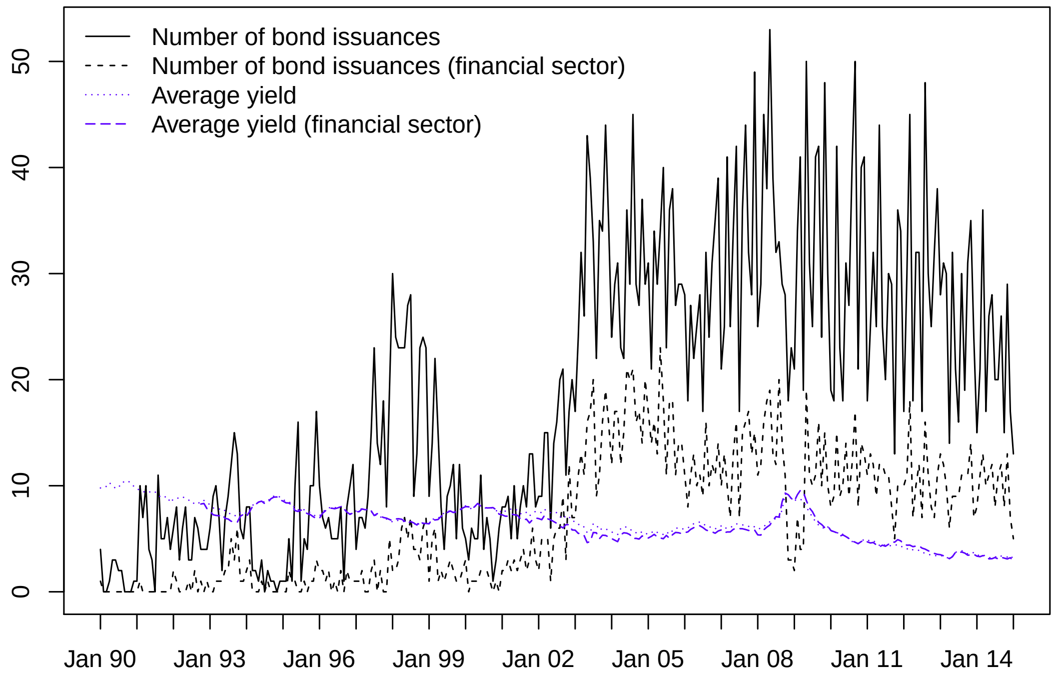
*Notes:* The dependent variable is 1 if a firm issues a bond in a particular month and zero otherwise. Yield is the firm-specific corporate bond yield and size is measured by assets. All specification include a measure of credit supply (leverage in the broker-dealer market) and the change in Federal Reserve Holdings of Treasury Notes as common factors. The first column uses all firms, the second column uses financial sector firms and the third column uses all other firms (excluding mining and agriculture). The last column reports the results when the common unobserved factors are omitted. t-statistics are shown in parenthesis.

**Table 3.12:** The effect of yields on bond issuance for US corporates in the post-crisis period by credit rating

| | All | | Financial | | Other | |
|---|---|---|---|---|---|---|
| | High | Low | High | Low | High | Low |
| Coefficient estimates | | | | | | |
| Yield | 0.031 | -0.085 | 0.052 | -0.045 | 0.016 | -0.205 |
| | (0.6) | (-1.174) | (0.599) | (-0.501) | (0.254) | (-2.049) |
| Size | -0.007 | -0.196 | -0.032 | -0.045 | -0.036 | -0.292 |
| | (-0.06) | (-4.37) | (-0.314) | (-1.487) | (-0.243) | (-4.875) |
| Marginal effects | | | | | | |
| Yield | 0.004 | -0.012 | 0.01 | -0.001 | 0.002 | -0.028 |
| Size | -0.004 | -0.025 | -0.01 | -0.006 | -0.006 | -0.036 |
| Observations | 160 | 160 | 40 | 25 | 112 | 106 |

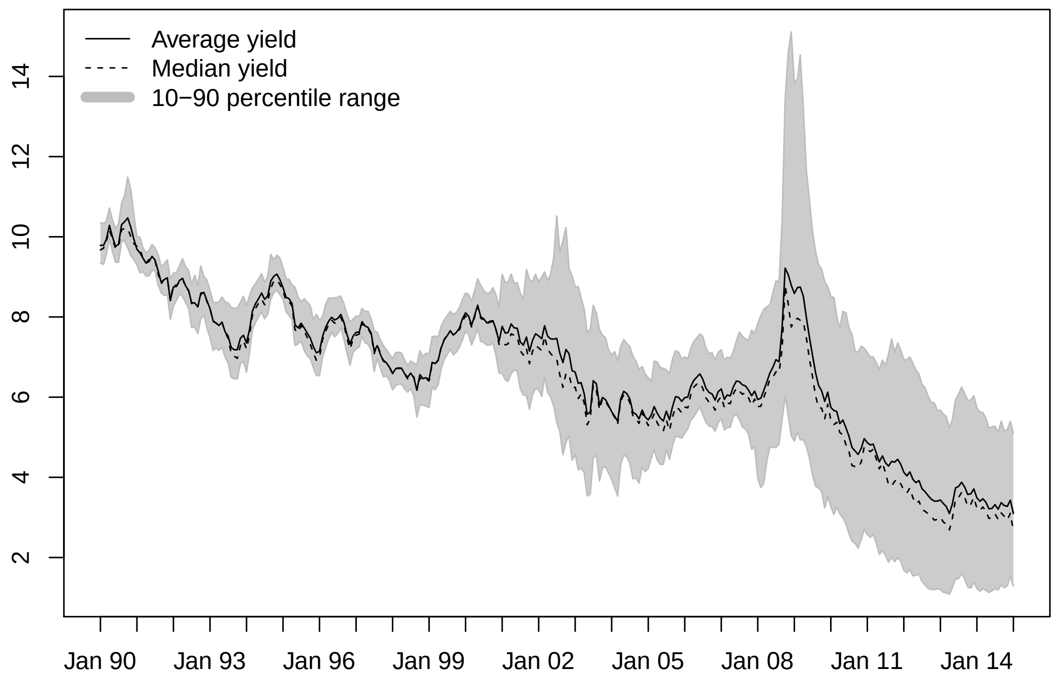
*Notes:* The dependent variable is 1 if a firm issues a bond in a particular month and zero otherwise. Yield the firm-specific corporate bond yield and size is measured by assets. All specification include a measure of credit supply (leverage in the broker-dealer market) and the change in Federal Reserve Holdings of Treasury Notes as common factors. Columns 1-2 use all firms, columns 3-4 use financial sector firms and columns 5-6 use all other firms (excluding mining and agriculture). Low (high) means that the issuer has a credit rating below (above) the sample median. t-statistics are shown in parenthesis.

**Figure 3.1:** Number of bond issuances per month and cross-sectional average of issuer-specific bond yields



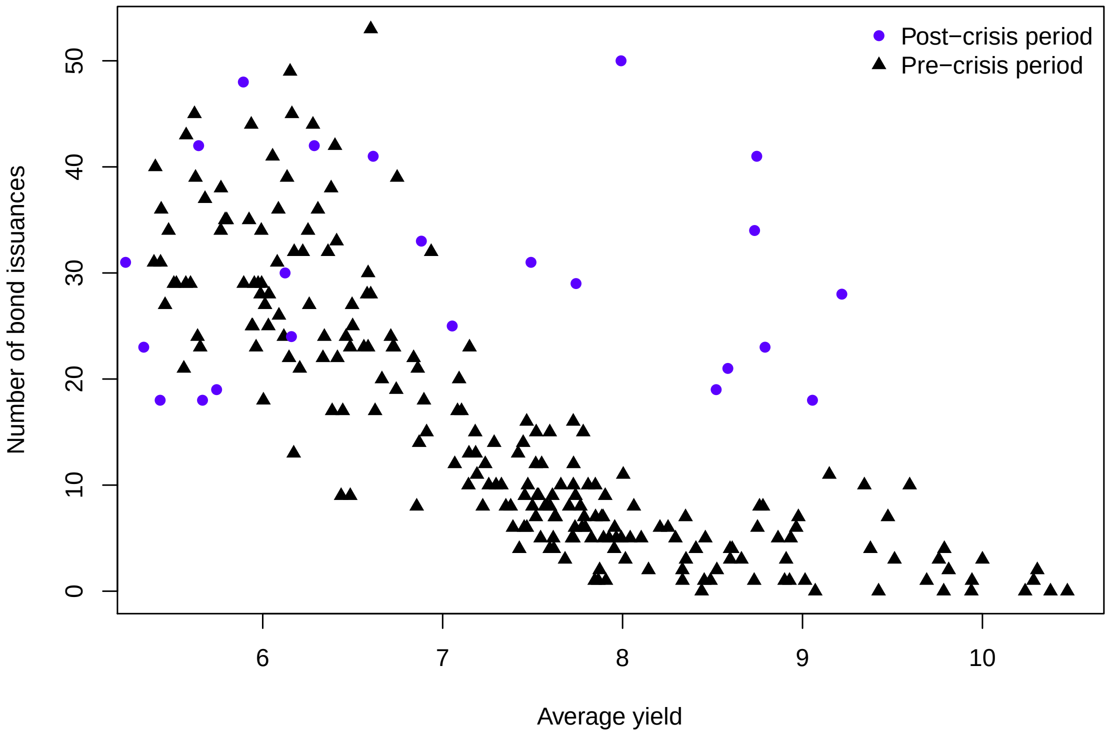

*Notes:* Data sources: Bloomberg, Datastream and own calculations.

**Figure 3.2:** Cross-sectional distribution of issuer-specific bond yields



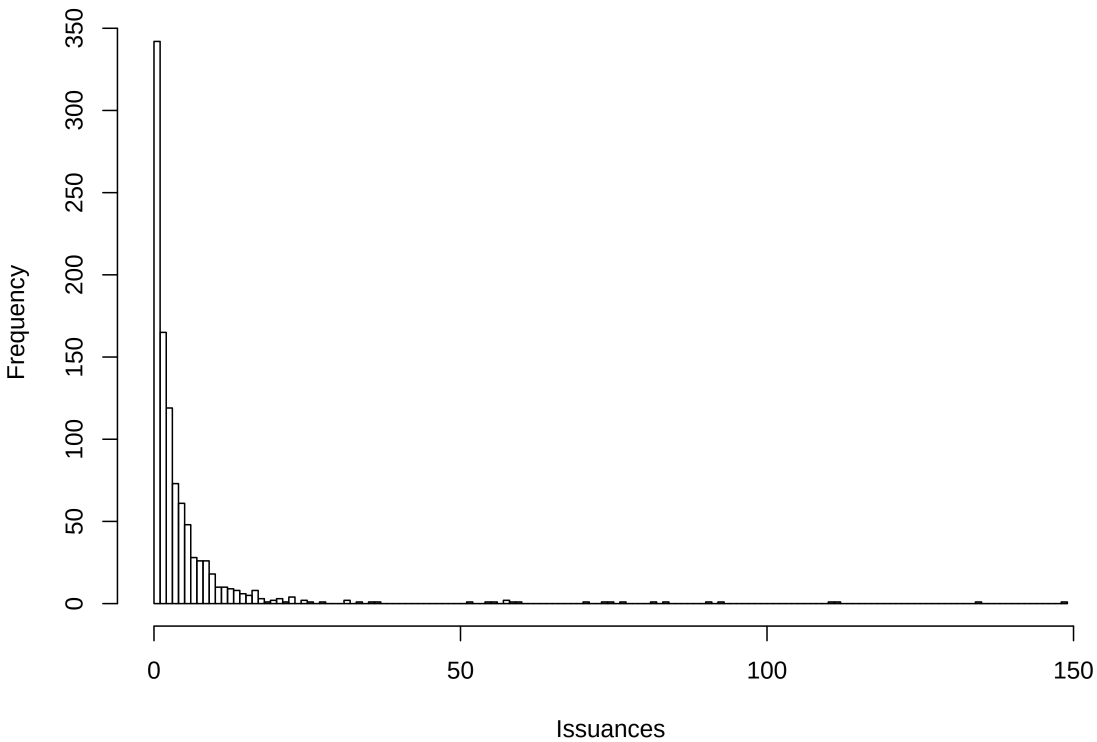

*Notes:* Data sources: Bloomberg, Datastream and own calculations.

**Figure 3.3:** Unconditional correlation between the number of bond issuances per month and cross-sectional average of issuer-specific bond yields




*Notes:* Data sources: Bloomberg, Datastream and own calculations.

**Figure 3.4:** Histogram for the number of issuances by firm




*Notes:* Data sources: Bloomberg, Datastream and own calculations.

# Bibliography


Adrian, T., P. Colla, and H. S. Shin (2012): "Which Financial Frictions? Parsing the Evidence from the Financial Crisis of 2007-9," *NBER Macroeconomics Annual 2012*, 27, 159–214.

Alizadeh, S., M. W. Brandt, and F. X. Diebold (2002): "Range-Based Estimation of Stochastic Volatility Models or Exchange Rate Dynamics are More Interesting Than You Think," *Journal of Finance*, 57, 1047–1091.

Amihud, Y. (2002): "Illiquidity and stock returns: cross-section and time-series effects," *Journal of Financial Markets*, 5, 31–56.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2010): "Parametric and Nonparametric Volatility Measurement," *Handbook of financial econometrics*, 1, 67–138.

Andrews, D. (1993): "An Introduction to Econometric Applications of Empirical Process Theory for Dependent Random Variables," *Econometric Reviews*, 12, 183–216.

Andrews, D. W. K. (2003): "Cross-section Regression with Common Shocks," Cowles Foundation Discussion Paper 1428, Cowles Foundation for Research in Economics, Yale University.

——— (2005): "Cross-Section Regression with Common Shocks," *Econometrica*, 73, 1551–1585.

Angrist, J. and J.-S. Pischke (2009): *Mostly harmless econometrics: An empiricist's companion*, Cambridge: Cambridge University Press.

Atak, A., O. Linton, and Z. Xiao (2011): "A semiparametric panel model for unbalanced data with application to climate change in the United Kingdom," *Journal of Econometrics*, 164, 92–115.

Baba, N. and M. Inada (2009): "Why do Japanese regional banks issue subordinated debts?" *Japan and the World Economy*, 21, 358–364.

Badoer, D. C. and C. M. James (2015): "The Determinants of Long-Term Corporate Debt Issuances," *The Journal of Finance*, forthcoming.

Bai, J. (2003): "Inferential Theory for Factor Models of Large Dimensions," *Econometrica*, 71, 135–171.

——— (2004): "Estimating cross-section common stochastic trends in nonstationary panel data," *Journal of Econometrics*, 122, 137–183.

——— (2009): "Panel Data Models With Interactive Fixed Effects," *Econometrica*, 77, 1229–1279.

Bai, J. and S. Ng (2002): "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70, 191–221.

Bailey, N., S. Holly, and M. H. Pesaran (2015): "A Two Stage Approach to Spatiotemporal Analysis with Strong and Weak Cross-Sectional Dependence," *Journal of Applied Econometrics*, forthcoming.

Baltagi, B. H., Q. Feng, and C. Kao (2016): "Estimation of Heterogeneous Panels with Structural Breaks," *Journal of Econometrics*, 191, 176–195.

Bartram, S. M., G. Brown, and R. M. Stulz (2012): "Why Are U.S. Stocks More Volatile?" *The Journal of Finance*, 67, 1329–1370.

Becker, B. and V. Ivashina (2014): "Cyclicality of credit supply: Firm level evidence," *Journal of Monetary Economics*, 62, 76–93.

Biais, B., D. Martimort, and J.-C. Rochet (2000): "Competing Mechanisms in a Common Value Environment," *Econometrica*, 68, 799–837.

Boehmer, B. and E. Boehmer (2003): "Trading your Neighbor's ETFs: Competition or Fragmentation?" *Journal of Banking & Finance*, 27, 1667–1703.

Boneva, L., O. Linton, and M. Vogt (2015): "The Effect of Fragmentation on Market Quality in the UK Equity Market," *Journal of Applied Econometrics*, forthcoming, –.

Bosq, D. (1998): *Nonparametric statistics for stochastic processes: estimation and prediction*, Berlin: Springer.

Bradley, R. C. (2005): "Basic Properties of Strong Mixing Conditions. A Survey and Some Open Questions," *Probability Surveys*, 2, 107–144.

Brealey, R., S. Myers, and F. Allen (2008): *Principles of Corporate Finance*, New York: McGraw-Hill/Irwin.

Brogaard, J., T. Hendershott, S. Hunt, T. Latza, L. Pedace, and C. Ysusi (2013): "High Frequency Trading and the Execution Costs of Long-Term Investors," *FSA occasional paper 43*.

Buti, S., B. Rindi, and I. M. Werner (2010): "Dynamic Dark Pool Trading Strategies in Limit Order Markets," *SSRN working paper*.

Campbell, J. Y., A. W. Lo, and A. C. MacKinlay (2012): *The Econometrics of Financial Markets*, Princeton University Press.

Chen, J., J. Gao, and D. Li (2012): "Semiparametric trending panel data models with cross-sectional dependence," *Journal of Econometrics*, 171, 71–85.

——— (2013a): "Estimation in Partially Linear Single-Index Panel Data Models With Fixed Effects," *Journal of Business & Economic Statistics*, 31, 315–330.

——— (2013b): "Estimation in Single-Index Panel Data Models with Heterogeneous Link Functions," *Econometric Reviews*, 32, 928–955.

Chen, M. (2014): "Estimation of Nonlinear Panel Models with Multiple Unobserved Effects," *mimo.*

Chen, M., I. Fernandez-Val, and M. Weidner (2014): "Nonlinear Panel Models with Interactive Effects," *arXiv:1412.5647 [stat]*, arXiv: 1412.5647.

Chen, X. (2007): "Large Sample Sieve Estimation of Semi-Nonparametric Models," *Handbook of Econometrics*, 6, Part B, 5549–5632.

——— (2013): "Penalized Sieve Estimation and Inference of Seminonparametric Dynamic Models: A Selective Review," in *Advances in Economics and Econometrics*, Cambridge University Press, vol. 3 of *Econometric Society Monographs.*

Chen, X., O. Linton, and I. Van Keilegom (2003): "Estimation of Semiparametric Models when the Criterion Function Is Not Smooth," *Econometrica*, 71, 1591–1608.

Chesini, G. (2012): "The Effects of Competition on the Fee Structures of the Major Stock Exchanges," *World Review of Business Research*, 2, 100–118.

Chudik, A. and M. H. Pesaran (2015a): "Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors," *Journal of Econometrics*, forthcoming.

——— (2015b): "Large panel data models with cross-sectional dependence: a survey," in *The Oxford Handbook of Panel Data*, ed. by B. H. Baltagi, Oxford Handbooks.

Chudik, A., M. H. Pesaran, and E. Tosetti (2011): "Weak and strong crosssection dependence and estimation of large panels," *Econometrics Journal*, 14, C45–C90.

Coakley, J., A.-M. Fuertes, and R. Smith (2002): "A Principal Components Approach to Cross-Section Dependence in Panels," *mimeo.*

——— (2006): "Unobserved heterogeneity in panel time series models," *Computational Statistics & Data Analysis*, 50, 2361–2380.

Conley, T. G. (1999): "GMM estimation with cross sectional dependence," *Journal of Econometrics*, 92, 1–45.

Connor, G., M. Hagmann, and O. Linton (2012): "Efficient Semiparametric Estimation of the FamaFrench Model and Extensions," *Econometrica*, 80, 713–754.

Connor, G. and R. A. Korajczyk (1988): "Risk and return in an equilibrium APT: Application of a new test methodology," *Journal of Financial Economics*, 21, 255–289.

——— (1993): "A Test for the Number of Factors in an Approximate Factor Model," *Journal of Finance*, 48, 1263–91.

Dahlhaus, R. (1997): "Fitting time series models to nonstationary processes," *The Annals of Statistics*, 25, 1–37.

De Jong, F., H. Degryse, and V. Van Kervel (2015): "The Impact of Dark Trading and Visible Fragmentation on Market Quality," *Review of Finance*, 19, 1587–1622.

Denis, D. J. and V. T. Mihov (2003): "The choice among bank debt, non-bank private debt, and public debt: evidence from new corporate borrowings," *Journal of Financial Economics*, 70, 3–28.

Dhyne, E., C. Fuss, M. H. Pesaran, and P. Sevestre (2011): "Lumpy Price Adjustments: A Microeconometric Analysis," *Journal of Business & Economic Statistics*, 29, 529–540.

Engle, R. F. and J. G. Rangel (2008): "The Spline-GARCH Model for Low-Frequency Volatility and Its Global Macroeconomic Causes," *Review of Financial Studies*, 21, 1187–1222.

European Commission (2011): "New Rules for More Efficient, Resilient and Transparent Financial Markets in Europe," *Press release*, IP/11/1219.

Fafchamps, M. and F. Gubert (2007): "The formation of risk sharing networks," *Journal of Development Economics*, 83, 326–350.

Fan, J. and I. Gijbels (1996): *Local polynomial modelling and its applications*, London: Chapman and Hall.

Farrant, K., M. Inkinen, M. Rutkowska, and K. Theodoridis (2013): "What can company data tell us about financing and investment decisions?" *Bank of England Quarterly Bulletin*, Q4, 361–370.

Fernandez-Val, I. and M. Weidner (2015): "Individual and time effects in nonlinear panel models with large N, T," *Journal of Econometrics*, forthcoming.

Firth, D. (1993): "Bias reduction of maximum likelihood estimates," *Biometrika*, 80, 27–38.

Foresight (2012): "The Future of Computer-based Trading in Financial Markets," *available at http://www.bis.gov.uk/foresight/our-work/projects/published-projects/computer-trading*.

Foucault, T. and A. J. Menkveld (2008): "Competition for Order Flow and Smart Order Routing Systems," *The Journal of Finance*, 63, 119–158.

Frank, M. Z. and V. K. Goyal (2008): "Chapter 12 - Trade-Off and Pecking Order Theories of Debt," in *Handbook of Empirical Corporate Finance*, ed. by B. E. Eckbo, San Diego: Elsevier, Handbooks in Finance, 135–202.

Gomber, P., G. Pujol, and A. Wranik (2012): "Best Execution Implementation and Broker Policies in Fragmented European Equity Markets," *International Review of Business Research Papers*, 8, 144–162.

Goyenko, R. Y., C. W. Holden, and C. A. Trzcinka (2009): "Do Liquidity Measures Measure Liquidity?" *Journal of Financial Economics*, 92, 153–181.

Grande, G., A. Levy, F. Panetta, and A. Zaghini (2011): "Public guarantees on bank bonds: effectiveness and distortions," *OECD Journal: Financial Market Trends*, 47–72.

Gresse, C. (2011): "Effects of the Competition between Multiple Trading Platforms on Market Liquidity: Evidence from the MiFID Experience," *SSRN working paper.*

——— (2012): "Market Fragmentation in Europe: Assessment and Prospects for Market Quality," *The Future of Computer Trading in Financial Markets: Foresight Driver Review 19, Government Office for Science.*

Haerdle, W. (1991): *Applied nonparametric regression*, Cambridge: Cambridge University Press.

Hansen, B. E. (2008): "Uniform Convergence Rates for Kernel Estimation with Dependent Data," *Econometric Theory*, 24, 726–748.

Harding, M. and C. Lamarche (2013): "Estimating and Testing a Quantile Regression Model with Interactive Effects," *Journal of Econometrics*, forthcoming, –.

Hart, O. D. and D. M. Kreps (1986): "Price Destabilizing Speculation," *Journal of Political Economy*, 94, 927–952.

Hendershott, T. (2011): "High Frequency Trading and Price Efficiency," *The Future of Computer Trading in Financial Markets: Foresight Driver Review 12, Government Office for Science.*

Henderson, D. J., R. J. Carroll, and Q. Li (2008): "Nonparametric estimation and testing of fixed effects panel data models," *Journal of Econometrics*, 144, 257–275.

Hengelbrock, J. and E. Theissen (2009): "Fourteen at One Blow: The Market Entry of Turquoise," *SSRN working paper.*

Hoderlein, S., E. Mammen, and K. Yu (2011): "Nonparametric models in binary choice fixed effects panel data," *Econometrics Journal*, 14, 351–367.

Hsiao, C. (2003): *Analysis of panel data*, Cambridge: Cambridge University Press.

Jenish, N. (2012): "Nonparametric spatial regression under near-epoch dependence," *Journal of Econometrics*, 167, 224–239.

Jolliffe, I. T. (2002): *Principal component analysis*, Berlin: Spriner.

Kapetanios, G., M. H. Pesaran, and T. Yamagata (2011): "Panels with Non-stationary Multifactor Error Structures," *Journal of Econometrics*, 160, 326–348.

Kervel, V. and Vincent (2015): "Competition for Order Flow with Fast and Slow Traders," *The Review of Financial Studies*, 28, 2094–2127.

Keys, B. J., T. Piskorski, A. Seru, and V. Yao (2014): "Mortgage Rates, Household Balance Sheets, and the Real Economy," Working Paper 20561, National Bureau of Economic Research.

Kneip, A., R. C. Sickles, and W. Song (2012): "A new panel data treatment for heterogeneity in time trends," *Econometric Theory*, 28, 590–628.

Koenker, R. (2005): *Quantile regression*, New York: Cambridge university press.

Kyriazidou, E. (1997): "Estimation of a Panel Data Sample Selection Model," *Econometrica*, 65, 1335–1364.

Lee, J. (2012): "Non-parametric methods under cross-sectional dependence," phd, The London School of Economics and Political Science (LSE).

Linton, O. (2012): "What has Happened to UK Equity Market Quality in the Last Decade? An Analysis of the Daily Data," *The Future of Computer Trading in Financial Markets: Foresight Driver Review 1, Government Office for Science.*

Linton, O., E. Maasoumi, and Y. Whang (2005): "Consistent Testing for Stochastic Dominance under General Sampling Schemes," *Review of Economic Studies*, 72, 735–765.

Linton, O. and J. P. Nielsen (1995): "A Kernel Method of Estimating Structured Nonparametric Regression Based on Marginal Integration," *Biometrika*, 82, 93–100.

Linton, O., J. P. Nielsen, and S. F. Nielsen (2009): "Non-parametric regression with a latent time series," *Econometrics Journal*, 12, 187–207.

Lo Duca, M., G. Nicoletti, and A. Vidal Martinez (2016): "Global corporate bond issuance: what role for US quantitative easing?" *Journal of International Money and Finance*, 60, 114–150.

Madhavan, A. (2012): "Exchange-traded funds, market structure and the flash crash," *SSRN working paper.*

Magnus, J. R. (1985): "On Differentiating Eigenvalues and Eigenvectors," *Econometric Theory*, 1, 179–191.

Mammen, E., O. Linton, and J. Nielsen (1999): "The existence and asymptotic properties of a backfitting projection algorithm under weak conditions," *The Annals of Statistics*, 27, 1443–1490.

Mammen, E., B. Stove, and D. Tjostheim (2009): "Nonparametric additive models for panels of time series," *Econometric Theory*, 25, 442–481.

Manski, C. (2008): *Identification for Prediction and Decision*, Harvard: Harvard University Press.

Masry, E. (1996): "Multivariate Local Polynomial Regression For Time Series: Uniform Strong Consistency And Rates," *Journal of Time Series Analysis*, 17, 571–599.

Mavroeidis, S., M. Plagborg-Mueller, and J. H. Stock (2014): "Empirical Evidence on Inflation Expectations in the New Keynesian Phillips Curve," *Journal of Economic Literature*, 52, 124–88.

Mizen, M. (2010): "Effective Spreads in European Equities," *TABB Group pinpoint.*

Mizen, P. and S. Tsoukas (2013): "What promotes greater use of the corporate bond market? A study of the issuance behaviour of firms in Asia," *Oxford Economic Papers*, 227–253.

Modigliani, F. and M. H. Miller (1958): "The Cost of Capital, Corporation Finance and the Theory of Investment," *The American Economic Review*, 48, 261–297.

Moon, H. and M. Weidner (2015): "Dynamic Linear Panel Regression Models with Interactive Fixed Effects," *Econometric Theory*, forthcoming.

Moon, H. R. and P. C. B. Phillips (1999): "Linear Regression Limit Theory for Nonstationary Panel Data," *Econometrica*, 67, 1057–1111.

Newey, W. K. (1994): "Series Estimation Of Regression Functionals," *Econometric Theory*, 10, 1–28.

Newey, W. K. and D. McFadden (1994): "Chapter 36 Large sample estimation and hypothesis testing," in *Handbook of Econometrics*, ed. by Robert F. Engle and Daniel L. McFadden, Elsevier, vol. Volume 4, 2111–2245.

Neyman, J. and E. L. Scott (1948): "Consistent Estimates Based on Partially Consistent Observations," *Econometrica*, 16, 1–32.

O'Hara, M. and M. Ye (2011): "Is Market Fragmentation Harming Market Quality?" *Journal of Financial Economics*, 100, 459–474.

Pagano, M. (1989): "Trading Volume and Asset Liquidity," *The Quarterly Journal of Economics*, 104, 255–274.

Parkinson, M. (1980): "The Extreme Value Method for Estimating the Variance of the Rate of Return," *The Journal of Business*, 53, 61–65.

Pesaran, M. H. (2006): "Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure," *Econometrica*, 74, 967–1012.

Pesaran, M. H. and E. Tosetti (2011): "Large panels with common factors and spatial correlation," *Journal of Econometrics*, 161, 182–202.

Porter, J. (1996): "Nonparametric regression estimation for a flexible panel data model," PhD Thesis, Department of Economics, MIT, Boston.

Qian, J. and L. Wang (2012): "Estimating semiparametric panel data models by marginal integration," *Journal of Econometrics*, 167, 483–493.

Riordan, R., A. Storkenmaier, and M. Wagener (2011): "Do Multilateral Trading Facilities Contribute to Market Quality?" *SSRN working paper*.

Robinson, P. M. (2011): "Asymptotic theory for nonparametric regression with spatial data," *Journal of Econometrics*, 165, 5–19.

Rogers, L. C. G. and S. E. Satchell (1991): "Estimating Variance From High, Low and Closing Prices," *The Annals of Applied Probability*, 1, 504–512.

ROSENBLATT, M. (1956): "A central limit theorem and a strong mixing condition," *Proceedings of the National Academy of Sciences of the United States of America*, 42, 43–47.

SECURITIES EXCHANGE COMMISSION (2013): "Equity market structure literature review, Part 1: Market fragmentation," *Working paper*.

SONG, M. (2013): "Asymptotic Theory for Dynamic Heterogeneous Panels with Cross-Sectional Dependence and Its Applications," *mimeo*.

STEIN, J. C. (1987): "Informational Externalities and Welfare-Reducing Speculation," *Journal of Political Economy*, 95, 11–23.

STONE, C. J. (1980): "Optimal Rates of Convergence for Nonparametric Estimators," *The Annals of Statistics*, 8, 1348–1360.

VOGT, M. (2012): "Nonparametric regression for locally stationary time series," *The Annals of Statistics*, 40, 2601–2633.

WAGENER, M. (2011): "Trading in European Equity Markets: Fragmentation and Market Quality," PhD thesis, Karlsruhe Institute of Technology, Karlsruhe.

WALD, A. (1949): "Note on the Consistency of the Maximum Likelihood Estimate," *The Annals of Mathematical Statistics*, 20, 595–601.

ZHU, M. AND A. GHODSI (2006): "Automatic dimensionality selection from the scree plot via the use of profile likelihood," *Computational Statistics & Data Analysis*, 51, 918–930.

# Appendices

# Appendix A

# A Semiparametric Model for Heterogeneous Panel Data with Fixed Effects

## A.1 Proofs

In this appendix, we derive the main results of our theory. In particular, we provide a detailed proof of Theorems 1.5.1 and 1.5.2, which characterize the asymptotic behaviour of our estimators. For the proof, we require a series of uniform convergence results which are derived in Appendix B. Throughout the appendix, the symbol $C$ is used to denote a universal real constant which may take a different value on each occurrence. Moreover, we let $I_h = [C_1 h, 1 - C_1 h]$ denote the interior of the support of the regressors $X_{it}$ and use $I_h^c = [0,1] \setminus I_h$ to denote the boundary region. Finally, we frequently make use of the shorthand $\kappa_0(x) = \int_{-x/h}^{(1-x)/h} K(\varphi) d\varphi$.

### Proof of Theorem 1.5.1

We restrict attention to the proof for the Nadaraya-Watson based estimators. The local linear case can be handled by similar arguments.

To start with, we list some auxiliary results needed to derive the statements (1.16) and (1.17) of Theorem 1.5.1. The proof of these results is postponed until the arguments for Theorem 1.5.1 are completed. The following uniform expansion of $\widehat{g}_k(x) - g_k(x)$ forms the basis of our arguments.

**Proposition A1.** *It holds that*

$$\widehat{g}_k(x) - g_k(x) = \sum_{i=1}^{n} \frac{\omega_{ki}}{\kappa_0(x) f_i(x)} \frac{1}{T} \sum_{t=1}^{T} K_h(X_{it} - x)\varepsilon_{it} + R_k(x), \tag{A.1}$$

*where the remainder satisfies* $\sup_{x \in I_h} |R_k(x)| = o_p(1/\sqrt{nTh})$ *and* $\sup_{x \in I_h^c} |R_k(x)| = O_p(h)$.

Using the uniform expansion of Proposition A1, we are able to derive the asymptotic properties of $\widehat{g}$. These are summarized in the next proposition.

**Proposition A2.** *It holds that*

$$\sup_{x \in I_h} \big\| \widehat{g}(x) - g(x) \big\| = O_p\Big( \sqrt{\frac{\log nT}{nTh}} \Big) \tag{A.2}$$

$$\sup_{x \in I_h^c} \big\| \widehat{g}(x) - g(x) \big\| = O_p(h). \tag{A.3}$$

*Moreover, for any fixed* $x \in (0,1)$,

$$\sqrt{nTh}(\widehat{g}(x) - g(x)) \xrightarrow{d} N(0, V(x)), \tag{A.4}$$

*where* $V(x) = (V_{k,l}(x))_{k,l=1,\ldots,K}$ *and* $V_{k,l}(x) = \|K\|_2^2 \lim_{n\to\infty}(n \sum_{i=1}^n \omega_{ki}\omega_{li} \frac{\sigma_i^2(x)}{f_i(x)})$ *with* $\sigma_i^2(x) = \mathbb{E}[\varepsilon_{it}^2 | X_{it} = x]$.

Proposition A1 can further be used to characterize the convergence behaviour of the matrices $\widehat{\Sigma}$.

**Proposition A3.** *It holds that*

$$\|\widehat{\Sigma} - \Sigma\| = o_p\Big(\frac{1}{\sqrt{nTh}}\Big). \tag{A.5}$$

Finally, Proposition A3 together with a Taylor expansion argument yields the following result.

**Proposition A4.** *It holds that*

$$\|\widehat{S} - S\| = o_p\Big(\frac{1}{\sqrt{nTh}}\Big) \tag{A.6}$$

$$\|\widehat{\lambda} - \lambda\| = o_p\Big(\frac{1}{\sqrt{nTh}}\Big) \tag{A.7}$$

*with* $\lambda = (\lambda_1, \dots, \lambda_K)^\intercal$ *and* $\widehat{\lambda} = (\widehat{\lambda}_1, \dots, \widehat{\lambda}_K)^\intercal$.

With the help of the above propositions, it is now straightforward to prove the statements (1.16) and (1.17) of Theorem 1.5.1. We start with the proof of (1.16): Recalling that the matrix of eigenvectors $S$ converges to a limit $S^*$ and using (A.2) together with (A.6), we arrive at

$$\begin{aligned}\sup_{x \in I_h} \|\widehat{\mu}(x) - \mu(x)\| &\le \|\widehat{S}^\intercal - S^\intercal\| \sup_{x \in I_h} \|\widehat{g}(x)\| \\ &\quad + \|S^\intercal\| \sup_{x \in I_h} \|\widehat{g}(x) - g(x)\| = O_p\Big(\sqrt{\frac{\log nT}{nTh}}\Big).\end{aligned}$$

Similarly, we obtain that

$$\begin{aligned}\sqrt{nTh}(\widehat{\mu}(x) - \mu(x)) &= \sqrt{nTh}(\widehat{S}^\intercal - S^\intercal)\widehat{g}(x) + S^\intercal \sqrt{nTh}(\widehat{g}(x) - g(x)) \\ &= S^\intercal \sqrt{nTh}(\widehat{g}(x) - g(x)) + o_p(1).\end{aligned}$$

Since $S$ converges to $S^*$, the normality result (A.4) implies that

$$S^\intercal \sqrt{nTh}(\widehat{g}(x) - g(x)) \xrightarrow{d} N(0, (S^*)^\intercal V(x) S^*),$$

which yields (1.17). $\square$

## Proof of Proposition A1

Let $\widehat{f}_i(x) = T^{-1} \sum_{t=1}^T K_h(X_{it} - x)$, $Y_{it}^{\text{fe}} = Y_{it} - \overline{Y}_i - \overline{Y}_t + \overline{\overline{Y}}$ and write

$$\widehat{g}_k(x) - g_k(x) = Q_{k,V}(x) + Q_{k,B}(x) + Q_{k,\gamma}(x) + Q_{k,\alpha} + Q_{k,\mu_0},$$

where

$$
\begin{aligned}
Q_{k,V}(x) &= \sum_{i=1}^{n} \omega_{ki} \frac{1}{T} \sum_{t=1}^{T} K_h(X_{it} - x)\varepsilon_{it} / \widehat{f}_i(x) \\
Q_{k,B}(x) &= \sum_{i=1}^{n} \omega_{ki} \frac{1}{T} \sum_{t=1}^{T} K_h(X_{it} - x)\big\{m_i(X_{it}) - m_i(x)\big\} / \widehat{f}_i(x) \\
Q_{k,\gamma}(x) &= \sum_{i=1}^{n} \omega_{ki} \frac{1}{T} \sum_{t=1}^{T} K_h(X_{it} - x)\big\{\mu_0 + \gamma_t - \overline{Y}_t\big\} / \widehat{f}_i(x) \\
Q_{k,\alpha} &= \sum_{i=1}^{n} \omega_{ki} \big\{\mu_0 + \alpha_i - \overline{Y}_i\big\} \\
Q_{k,\mu_0} &= \Big(\sum_{i=1}^{n} \omega_{ki}\Big) \big\{\overline{\overline{Y}} - \mu_0\big\}.
\end{aligned}
$$

In what follows, we analyze these five terms one after the other.

(i) It holds that

$$
Q_{k,V}(x) = \sum_{i=1}^{n} \omega_{ki} \frac{1}{T} \sum_{t=1}^{T} K_h(X_{it} - x)\varepsilon_{it} / \kappa_0(x) f_i(x) + R_{k,V}(x),
$$

where the remainder term is given by

$$
R_{k,V}(x) = \sum_{m=1}^{M} R_{k,V}^{(m)}(x) + R_{k,V}^{(M+1)}(x) + R_{k,V}^{(B)}(x)
$$

with

$$
R_{k,V}^{(m)}(x) = \sum_{i=1}^{n} \omega_{ki} \Big( \frac{(\mathbb{E}[\widehat{f}_i(x)] - \widehat{f}_i(x))^m}{\mathbb{E}[\widehat{f}_i(x)]^{m+1}} \Big) \Big( \frac{1}{T} \sum_{t=1}^{T} K_h(X_{it} - x)\varepsilon_{it} \Big)
$$

for $m = 1, \dots, M$,

$$
R_{k,V}^{(M+1)}(x) = \sum_{i=1}^{n} \omega_{ki} \Big( \frac{(\mathbb{E}[\widehat{f}_i(x)] - \widehat{f}_i(x))^{M+1}}{\mathbb{E}[\widehat{f}_i(x)]^{M+1} \widehat{f}_i(x)} \Big) \Big( \frac{1}{T} \sum_{t=1}^{T} K_h(X_{it} - x)\varepsilon_{it} \Big)
$$

and

$$
R_{k,V}^{(B)}(x) = \sum_{i=1}^{n} \omega_{ki} \Big( \frac{\kappa_0(x) f_i(x) - \mathbb{E}[\widehat{f}_i(x)]}{\kappa_0(x) f_i(x) \mathbb{E}[\widehat{f}_i(x)]} \Big) \Big( \frac{1}{T} \sum_{t=1}^{T} K_h(X_{it} - x)\varepsilon_{it} \Big).
$$

The remainder term has the property that

$$
\sup_{x \in I_h} \big| R_{k,V}(x) \big| = o_p\Big( \frac{1}{\sqrt{nTh}} \Big) \tag{A.8}
$$

$$
\sup_{x \in I_h^c} \big| R_{k,V}(x) \big| = O_p(h). \tag{A.9}
$$

We first derive (A.8): To start with, straightforward calculations yield that $\max_{1 \le i \le n}$

$\sup_{x\in I_h} |\kappa_0(x) f_i(x) - \mathbb{E}[\widehat{f}_i(x)]| = O_p(h^2)$. Together with Lemma B1 in Appendix B, this directly implies that $\sup_{x\in I_h} |R_{k,V}^{(B)}(x)| = o_p(1/\sqrt{nTh})$. Moreover, by Lemma B3, it holds that $\sup_{x\in I_h} |R_{k,V}^{(m)}(x)| = o_p(1/\sqrt{nTh})$ for $m = 1, \ldots, M$. Finally, if $M$ is chosen sufficiently large, then an application of Lemma B1 immediately shows that $\sup_{x\in I_h} |R_{k,V}^{(M+1)}(x)| = o_p(1/\sqrt{nTh})$ as well. (A.9) follows by analogous arguments.

(ii) We next show that

$$\sup_{x\in I_h} |Q_{k,B}(x)| = o_p\Big(\frac{1}{\sqrt{nTh}}\Big)$$
$$\sup_{x\in I_h^c} |Q_{k,B}(x)| = O_p(h).$$

To see this, decompose $Q_{k,B}(x)$ into the following two components:

$$Q_{k,B}(x) = Q_{k,B}^{(1)}(x) + Q_{k,B}^{(2)}(x)$$

with

$$\begin{aligned} Q_{k,B}^{(1)}(x) = \sum_{i=1}^{n} \omega_{ki} \frac{1}{T} \sum_{t=1}^{T} \Big( & K_h(X_{it}-x)\big\{m_i(X_{it}) - m_i(x)\big\} \\ & - \mathbb{E}\big[K_h(X_{it}-x)\big\{m_i(X_{it}) - m_i(x)\big\}\big]\Big) \Big/ \widehat{f}_i(x) \end{aligned}$$

$$Q_{k,B}^{(2)}(x) = \sum_{i=1}^{n} \omega_{ki} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}\big[K_h(X_{it}-x)\big\{m_i(X_{it}) - m_i(x)\big\}\big] \big/ \widehat{f}_i(x).$$

Exploiting the smoothness conditions on the functions $m_i$ and $f_i$ in a standard way, the term $Q_{k,B}^{(2)}(x)$ can be shown to satisfy $\sup_{x\in I_h} |Q_{k,B}^{(2)}(x)| = O_p(h^2) = o_p(1/\sqrt{nTh})$ and $\sup_{x\in I_h^c} |Q_{k,B}^{(2)}(x)| = O_p(h)$. Moreover, $Q_{k,B}^{(1)}(x) = Q_{k,B}^{(1,a)}(x) + Q_{k,B}^{(1,b)}(x)$ with

$$\begin{aligned} Q_{k,B}^{(1,a)}(x) = \sum_{i=1}^{n} \omega_{ki} \frac{1}{T} \sum_{t=1}^{T} \Big( & K_h(X_{it}-x)\big\{m_i(X_{it}) - m_i(x)\big\} \\ & - \mathbb{E}\big[K_h(X_{it}-x)\big\{m_i(X_{it}) - m_i(x)\big\}\big]\Big) \Big/ \kappa_0(x) f_i(x) \end{aligned}$$

$$\begin{aligned} Q_{k,B}^{(1,b)}(x) = \sum_{i=1}^{n} \omega_{ki} \Big( \frac{\kappa_0(x) f_i(x) - \widehat{f}_i(x)}{\kappa_0(x) f_i(x) \widehat{f}_i(x)} \Big) \frac{1}{T} \sum_{t=1}^{T} \Big( & K_h(X_{it}-x)\big\{m_i(X_{it}) - m_i(x)\big\} \\ & - \mathbb{E}\big[K_h(X_{it}-x)\big\{m_i(X_{it}) - m_i(x)\big\}\big]\Big). \end{aligned}$$

Using the proof strategy of Lemma B2, the term $Q_{k,B}^{(1,a)}(x)$ can be shown to be of the order $O_p(h\sqrt{\log nT/nTh}) = o_p(1/\sqrt{nTh})$ uniformly for $x \in [0,1]$. Moreover, applying Lemma B1, it is straightforward to see that $\sup_{x\in[0,1]} |Q_{k,B}^{(1,b)}(x)| = o_p(1/\sqrt{nTh})$ as well.

(iii) We now turn to the analysis of $Q_{k,\gamma}(x)$. In particular, we show that

$$\sup_{x\in[0,1]} |Q_{k,\gamma}(x)| = o_p\Big(\frac{1}{\sqrt{nTh}}\Big).$$

To do so, first note that

$$Q_{k,\gamma}(x) = -\sum_{i=1}^n \omega_{ki}\frac{1}{T}\sum_{t=1}^T K_h(X_{it}-x)\Big\{\frac{1}{n}\sum_{j=1}^n (m_j(X_{jt})+\varepsilon_{jt})\Big\}\Big/\widehat{f}_i(x).$$

This expression can be decomposed as follows: $Q_{k,\gamma}(x) = Q_{k,\gamma}^{(1)}(x)+Q_{k,\gamma}^{(2)}(x)+Q_{k,\gamma}^{(3)}(x)$, where

$$\begin{aligned}
Q_{k,\gamma}^{(1)}(x) &= -\sum_{i=1}^n \omega_{ki}\frac{1}{T}\sum_{t=1}^T K_h(X_{it}-x)\Big\{\frac{1}{n}\sum_{j=1}^n (m_j(X_{jt})+\varepsilon_{jt})\Big\}\Big/\kappa_0(x)f_i(x) \\
Q_{k,\gamma}^{(2)}(x) &= -\sum_{i=1}^n \omega_{ki}\frac{1}{T}\sum_{t=1}^T K_h(X_{it}-x)\Big\{\frac{1}{n}(m_i(X_{it})+\varepsilon_{it})\Big\}\Big(\frac{1}{\widehat{f}_i(x)}-\frac{1}{\kappa_0(x)f_i(x)}\Big) \\
Q_{k,\gamma}^{(3)}(x) &= -\sum_{i=1}^n \omega_{ki}\frac{1}{T}\sum_{t=1}^T K_h(X_{it}-x)\Big\{\frac{1}{n}\sum_{\substack{j=1\\ j\neq i}}^n (m_j(X_{jt})+\varepsilon_{jt})\Big\}\Big(\frac{1}{\widehat{f}_i(x)}-\frac{1}{\kappa_0(x)f_i(x)}\Big).
\end{aligned}$$

To analyze the term $Q_{k,\gamma}^{(1)}(x)$, we further split it up into two components: $Q_{k,\gamma}^{(1)}(x) = Q_{k,\gamma}^{(1,a)}(x) + Q_{k,\gamma}^{(1,b)}(x)$, where

$$\begin{aligned}
Q_{k,\gamma}^{(1,a)}(x) &= -\frac{1}{T}\sum_{t=1}^T \Big(\sum_{i=1}^n \frac{\omega_{ki}}{\kappa_0(x)f_i(x)}(K_h(X_{it}-x)-\mathbb{E}[K_h(X_{it}-x)])\Big) \\
&\qquad\qquad\qquad\qquad \times \Big\{\frac{1}{n}\sum_{j=1}^n (m_j(X_{jt})+\varepsilon_{jt})\Big\} \\
Q_{k,\gamma}^{(1,b)}(x) &= -\sum_{i=1}^n \frac{\omega_{ki}}{\kappa_0(x)f_i(x)}\Big(\frac{1}{nT}\sum_{j=1}^n\sum_{t=1}^T \mathbb{E}[K_h(X_{it}-x)](m_j(X_{jt})+\varepsilon_{jt})\Big).
\end{aligned}$$

The term $Q_{k,\gamma}^{(1,a)}(x)$ can be handled by similar techniques as applied in Lemma B3. The details are summarized in Lemma B4 which yields that $\sup_{x\in[0,1]}|Q_{k,\gamma}^{(1,a)}(x)| = o_p(1/\sqrt{nTh})$. Moreover, it is straightforward to verify that $\sup_{x\in[0,1]}|Q_{k,\gamma}^{(1,b)}(x)| = O_p(1/\sqrt{nT})$. Turning to the expression $Q_{k,\gamma}^{(2)}(x)$, we can easily see with the help of Lemma B1 that $\sup_{x\in[0,1]}|Q_{k,\gamma}^{(2)}(x)| = o_p(1/\sqrt{nTh})$. To prove that $\sup_{x\in[0,1]}|Q_{k,\gamma}^{(3)}(x)| = o_p(1/\sqrt{nTh})$, some rather involved arguments are needed which are presented in Lemma B5. Setting $\widehat{\phi}_i(x) = (\widehat{f}_i(x))^{-1}-(\kappa_0(x)f_i(x))^{-1}$ in this lemma yields the result.

Finally, it is trivial to see that $Q_{k,\alpha} = O_p(1/\sqrt{nT})$ as well as $Q_{k,\mu_0} = O_p(1/\sqrt{nT})$. Together with (i)–(iii), this yields the expansion (A.1). $\square$

## Proof of Proposition A2

The proof easily follows with the help of the uniform expansion from Proposition A1. The latter says that

$$\widehat{g}_k(x) - g_k(x) = W_{k,V}(x) + R_k(x),$$

where

$$W_{k,V}(x) = \sum_{i=1}^{n} \omega_{ki} \frac{1}{T} \sum_{t=1}^{T} K_h(X_{it} - x)\varepsilon_{it} / \kappa_0(x) f_i(x)$$

and the remainder term $R_k(x)$ satisfies $\sup_{x \in I_h} |R_k(x)| = o_p(1/\sqrt{nTh})$ as well as $\sup_{x \in I_h^c} |R_k(x)| = O_p(h)$. Applying Lemma B2 to $W_{k,V}(x)$, we immediately obtain that $\sup_{x \in [0,1]} |W_{k,V}(x)| = O_p(\sqrt{\log nT/nTh})$. This yields the uniform convergence results (A.2) and (A.3). Furthermore, standard arguments show that

$$\sqrt{nTh} W_{k,V}(x) \xrightarrow{d} N\Big(0, \|K\|_2^2 \lim_{n \to \infty} n \sum_{i=1}^{n} \omega_{ki}^2 \frac{\sigma_i^2(x)}{f_i(x)}\Big).$$

From this, the normality result (A.4) easily follows. $\square$

## Proof of Proposition A3

It holds that

$$\begin{aligned}
\widehat{\Sigma}_{kl} - \Sigma_{kl} &= \int \widehat{g}_k(x)\widehat{g}_l(x)w(x)dx - \int g_k(x)g_l(x)w(x)dx \\
&= \int \big[\widehat{g}_k(x) - g_k(x)\big]\widehat{g}_l(x)w(x)dx + \int g_k(x)\big[\widehat{g}_l(x) - g_l(x)\big]w(x)dx \\
&= \int \big[\widehat{g}_k(x) - g_k(x)\big]g_l(x)w(x)dx + \int g_k(x)\big[\widehat{g}_l(x) - g_l(x)\big]w(x)dx \\
&\quad + o_p\Big(\frac{1}{\sqrt{nTh}}\Big),
\end{aligned}$$

where the last equality follows by Proposition A2. Using the uniform expansion of Proposition A1, we obtain

$$\int \big[\widehat{g}_k(x) - g_k(x)\big]g_l(x)w(x)dx = J_V + R$$

with

$$J_V = \sum_{i=1}^{n} \omega_{ki} \frac{1}{T} \sum_{t=1}^{T} \Big( \int K_h(X_{it} - x) g_l(x) (\kappa_0(x) f_i(x))^{-1} w(x) dx \Big) \varepsilon_{it}$$

and $R = \int g_l(x)R_k(x)w(x)dx$. As $\sup_{x \in I_h} |R_k(x)| = o_p(1/\sqrt{nTh})$ and $\sup_{x \in I_h^c} |R_k(x)| = O_p(h)$, we have that $R = o_p(1/\sqrt{nTh})$. Moreover, applying Chebychev's inequality and exploiting the mixing conditions on the data with the help of Davydov's inequality (see Corollary 1.1 in Bosq (1998)), it is not difficult to see that $J_V = o_p(1/\sqrt{nTh})$. This completes the proof. $\square$

## Proof of Proposition A4

Let $v(A) = \text{vec}(A)$ be the vectorized representation of a $K \times K$ matrix $A$. There are fixed vector-valued functions $f_k(\cdot)$ and scalar functions $\psi_k(\cdot)$ with first and second derivatives existing and being continuous in a neighbourhood of $v(\Sigma^*)$ such that

$$s_k = f_k(v(\Sigma)) \quad \text{and} \quad \lambda_k = \psi_k(v(\Sigma))$$
$$\widehat{s}_k = f_k(v(\widehat{\Sigma})) \quad \text{and} \quad \widehat{\lambda}_k = \psi_k(v(\widehat{\Sigma}))$$

(cp. Magnus (1985)). In what follows, we show that $\|\widehat{s}_k - s_k\| = o_p(1/\sqrt{nTh})$ for all $k = 1, \ldots, K$, which immediately yields (A.6). The result (A.7) for the estimates of the eigenvalues follows by exactly the same argument. From Proposition A3, we know that

$$\|v(\widehat{\Sigma}) - v(\Sigma)\| = o_p\Big(\frac{1}{\sqrt{nTh}}\Big).$$

As $f_k$ is continuously differentiable in a neighbourhood of $v(\Sigma^*)$, a first-order Taylor expansion yields

$$\widehat{s}_k - s_k = f_k(v(\widehat{\Sigma})) - f_k(v(\Sigma)) = f_k'(\xi)\big[v(\widehat{\Sigma}) - v(\Sigma)\big]$$

with $\xi$ being an intermediate point between $v(\widehat{\Sigma})$ and $v(\Sigma)$. Since $f_k'(\xi) - f_k'(v(\Sigma^*)) = o_p(1)$, we immediately arrive at

$$\|\widehat{s}_k - s_k\| = o_p\Big(\frac{1}{\sqrt{nTh}}\Big). \qquad \square$$

## Proof of Theorem 1.5.2

We again restrict attention to the Nadaraya-Watson based case, the arguments for the local linear case being essentially the same. Write

$$\sqrt{T}(\widehat{\beta}_i - \beta_i) = \sqrt{T}(\widehat{\beta}_i - \widetilde{\beta}_i) + \sqrt{T}(\widetilde{\beta}_i - \beta_i),$$

where $\widetilde{\beta}_i$ is the infeasible parameter estimator defined in (1.11). In what follows, we analyze the two terms on the right-hand side separately.

(i) First consider the term $\sqrt{T}(\widehat{\beta}_i - \widetilde{\beta}_i)$. It holds that

$$\begin{aligned}
&\sqrt{T}(\widehat{\beta}_i - \widetilde{\beta}_i) \\
&= \Big(\frac{1}{T}\sum_{t=1}^{T}\pi(X_{it})\widehat{\mu}(X_{it})\widehat{\mu}(X_{it})^\intercal\Big)^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\pi(X_{it})\big\{\widehat{\mu}(X_{it})-\mu(X_{it})\big\}Y_{it}^{\text{fe}} \\
&\quad + \Big\{\Big(\frac{1}{T}\sum_{t=1}^{T}\pi(X_{it})\widehat{\mu}(X_{it})\widehat{\mu}(X_{it})^\intercal\Big)^{-1} - \Big(\frac{1}{T}\sum_{t=1}^{T}\pi(X_{it})\mu(X_{it})\mu(X_{it})^\intercal\Big)^{-1}\Big\} \\
&\qquad \times \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\pi(X_{it})\mu(X_{it})Y_{it}^{\text{fe}}.
\end{aligned}$$

Here,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\pi(X_{it})\mu(X_{it})Y_{it}^{\text{fe}} = L_1 + L_2 + L_3 + L_4$$

with

$$\begin{aligned}
L_1 &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\pi(X_{it})\mu(X_{it})\varepsilon_{it} \\
L_2 &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\pi(X_{it})\mu(X_{it})m_i(X_{it}) \\
L_3 &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\pi(X_{it})\mu(X_{it})\big(\mu_0 + \gamma_t - \overline{Y}_t\big) \\
L_4 &= \Big(\frac{1}{T}\sum_{t=1}^{T}\pi(X_{it})\mu(X_{it})\Big)\sqrt{T}\big(\alpha_i - \overline{Y}_i + \overline{\overline{Y}}\big).
\end{aligned}$$

It is straightforward to see that $L_1 = O_p(1)$, $L_2 = O_p(\sqrt{T})$, $L_3 = o_p(1)$ and $L_4 = O_p(1)$. Hence,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\pi(X_{it})\mu(X_{it})Y_{it}^{\text{fe}} = O_p(\sqrt{T}). \tag{A.10}$$

As $\sup_{x\in I_h}\|\widehat{\mu}(x)-\mu(x)\| = O_p(\sqrt{\log nT/nTh}) = o_p(1/\sqrt{T})$, we further obtain that

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}\pi(X_{it})\widehat{\mu}(X_{it})\widehat{\mu}(X_{it})^\intercal - \frac{1}{T}\sum_{t=1}^{T}\pi(X_{it})\mu(X_{it})\mu(X_{it})^\intercal \\
= O_p\Big(\sqrt{\frac{\log nT}{nTh}}\Big) = o_p\Big(\frac{1}{\sqrt{T}}\Big)
\end{aligned} \tag{A.11}$$

as well as

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\pi(X_{it})\big\{\widehat{\mu}(X_{it})-\mu(X_{it})\big\}Y_{it}^{\text{fe}} = o_p(1). \tag{A.12}$$

Combining (A.10)–(A.12) yields $\sqrt{T}(\widehat{\beta}_i - \widetilde{\beta}_i) = o_p(1)$.

(ii) We next turn to $\sqrt{T}(\widetilde{\beta}_i - \beta_i)$. Write

$$\sqrt{T}(\widetilde{\beta}_i - \beta_i) = \Big(\frac{1}{T}\sum_{t=1}^{T} \pi(X_{it})\mu(X_{it})\mu(X_{it})^\intercal\Big)^{-1}(L_1 + L_3 + L_4)$$

with $L_1$, $L_3$ and $L_4$ introduced above. Since $L_3 = o_p(1)$ and $T^{-1}\sum_{t=1}^{T}\pi(X_{it})\mu(X_{it}) \xrightarrow{P} \mathbb{E}[\pi(X_{it})\mu(X_{it})]$, we can rewrite $L_4$ as

$$L_4 = -\mathbb{E}[\pi(X_{it})\mu(X_{it})]\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(m_i(X_{it}) + \varepsilon_{it}) + o_p(1).$$

This yields that

$$L_1 + L_3 + L_4 = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\chi_{it} + o_p(1),$$

where $\chi_{it} = (\pi(X_{it})\mu(X_{it}) - \mathbb{E}[\pi(X_{it})\mu(X_{it})])\varepsilon_{it} - \mathbb{E}[\pi(X_{it})\mu(X_{it})]m_i(X_{it})$. Applying a central limit theorem, we now arrive at

$$\sqrt{T}(\widetilde{\beta}_i - \beta_i) \xrightarrow{d} N(0, \Gamma_i^{-1}\Psi_i(\Gamma_i^{-1})^\intercal),$$

where the matrices $\Gamma_i$ and $\Psi_i$ are given by $\Gamma_i = \mathbb{E}[\pi(X_{it})\mu(X_{it})\mu(X_{it})^\intercal]$ and $\Psi_i = \sum_{l=-\infty}^{\infty}\mathrm{Cov}(\chi_{i0}, \chi_{il})$. $\square$

## Proof of Theorem 1.6.2

The same arguments as for the proof of Proposition A3 show that

$$\|\widetilde{\Sigma} - \overline{\Sigma}\| = o_p\Big(\frac{1}{\sqrt{nTh}}\Big).$$

Moreover, letting $\overline{\lambda}_1 \geq \ldots \geq \overline{\lambda}_{\overline{K}}$ be the eigenvalues of the matrix $\overline{\Sigma}$ and $\widetilde{\lambda}_1 \geq \ldots \geq \widetilde{\lambda}_{\overline{K}}$ the eigenvalues of $\widetilde{\Sigma}$, we have that

$$\widetilde{\lambda}_k = \int \widetilde{\mu}_k^2(x)w(x)dx$$

and $\overline{\lambda}_k = 0$ for $k = K+1, \ldots, \overline{K}$. Finally, note that the mapping of symmetric matrices to their eigenvalues is Lipschitz continuous. In particular, let $A$ and $B$ be any real symmetric $\overline{K} \times \overline{K}$ matrices and let $\lambda_1(A) \geq \lambda_2(A) \geq \ldots \geq \lambda_{\overline{K}}(A)$ and $\lambda_1(B) \geq \lambda_2(B) \geq \ldots \geq \lambda_{\overline{K}}(B)$ be the corresponding eigenvalues. Then there exists a constant $L$ independent of $A$ and $B$ such that

$$|\lambda_k(A) - \lambda_k(B)| \leq L\|A - B\|.$$

Combining the above remarks, we arrive at

$$\int \widetilde{\mu}_k^2(x)w(x)dx = \widetilde{\lambda}_k = |\widetilde{\lambda}_k - \overline{\lambda}_k| \le L\|\widetilde{\Sigma} - \overline{\Sigma}\| = o_p\Big(\frac{1}{\sqrt{nTh}}\Big).$$

for all $k = K+1, \dots, \overline{K}$. □

## A.2 Supplementary results on uniform convergence

In this appendix, we list some lemmas on uniform convergence which are needed to derive the main theorems. To prove the lemmas, we use a covering argument together with an exponential inequality, thus following the common strategy to be found for example in Bosq (1998), Masry (1996) or Hansen (2008). For the proof of Lemmas B1 and B2, these standard arguments have to be modified only slightly. For the proof of Lemmas B3–B5 in contrast, some rather intricate and non-standard arguments are needed to get the overall strategy to work.

We formulate the results for a general array $\{(X_{it}, Z_{it})\} = \{(X_{it}, Z_{it}),\ i = 1, \dots, n,\ t = 1, \dots, T\}$ which satisfies the following conditions:

(A1') The data $\{(X_{it}, Z_{it})\}$ are independent across $i$. Moreover, they are strictly stationary and strongly mixing in the time direction. Let $\alpha_i(k)$ for $k = 1, 2, \dots$ be the mixing coefficients of the time series $\{(X_{it}, Z_{it}), t = 1, \dots, T\}$ of the $i$-th individual. It holds that $\alpha_i(k) \le \alpha(k)$ for all $i = 1, \dots, n$, where the coefficients $\alpha(k)$ decay exponentially fast to zero as $k \to \infty$.

(A4') For some $\theta > 5$ and for all $l \in \mathbb{Z}$,

$$\max_{1\le i\le n} \sup_{x\in[0,1]} \mathbb{E}\big[|Z_{it}|^\theta \big| X_{it} = x\big] \le C < \infty$$
$$\max_{1\le i\le n} \sup_{x,x'\in[0,1]} \mathbb{E}\big[|Z_{it}|\big| X_{it} = x, X_{it+l} = x'\big] \le C < \infty$$
$$\max_{1\le i\le n} \sup_{x,x'\in[0,1]} \mathbb{E}\big[|Z_{it}Z_{it+l}|\big| X_{it} = x, X_{it+l} = x'\big] \le C < \infty,$$

where $C$ is a sufficiently large constant independent of $l$.

In addition, we suppose that the variables $X_{it}$ and $(X_{it}, X_{it+l})$ have densities $f_i$ and $f_{i;l}$ which satisfy (A2) and that the kernel $K$ and the dimensions $n$ and $T$ fulfill (A5)–(A7).

Throughout the appendix, we assume that the above conditions are satisfied. We now formulate the various results:

**Lemma B1.** *For kernel averages $\Psi_i(x)$ of the form*

$$\Psi_i(x) = \frac{1}{T}\sum_{t=1}^{T} K_h(X_{it} - x) Z_{it},$$

*it holds that*

$$\max_{1\le i\le n} \sup_{x\in[0,1]} \big|\Psi_i(x) - \mathbb{E}[\Psi_i(x)]\big| = o_p(1). \tag{A.13}$$

*If the variables $Z_{it}$ are bounded, i.e., if $|Z_{it}| \le C$ for some constant $C$ independent of $i$ and $t$, then we even have that*

$$\max_{1\le i\le n} \sup_{x\in[0,1]} \big|\Psi_i(x) - \mathbb{E}[\Psi_i(x)]\big| = O_p\Big(\sqrt{\frac{\log T}{Th}}\Big). \tag{A.14}$$

**Proof of Lemma B1.** The proof proceeds by slightly modifying standard arguments to derive uniform convergence rates for kernel estimators. We are thus content with giving some remarks on the necessary modifications.

We start with the proof of (A.14). Write

$$\mathbb{P}\Big( \max_{1\le i\le n} \sup_{x\in[0,1]} \big|\Psi_i(x) - \mathbb{E}[\Psi_i(x)]\big| > Ca_T\Big) \le \sum_{i=1}^{n} \mathbb{P}\Big( \sup_{x\in[0,1]} \big|\Psi_i(x) - \mathbb{E}[\Psi_i(x)]\big| > Ca_T\Big)$$

with $a_T = \sqrt{\log T/Th}$. Going along the lines of the standard proving strategy, the probabilities on the right-hand side can be bounded by a null sequence $\{c_T\}$ which does not depend on $i$. Under our conditions, this sequence can be chosen such that $\{nc_T\}$ is a null sequence as well. This yields the result.

We now turn to (A.13). As the variables $Z_{it}$ are not bounded, we have to replace them by truncated versions $Z_{it}^{\le} = Z_{it} I(|Z_{it}| \le \tau_{n,T})$ in a first step. Since we maximize over $i$, the truncation sequence $\tau_{n,T}$ must be chosen to go to infinity much faster than in the standard case where $i$ is fixed. In particular, we take $\tau_{n,T} = (nT)^{1/(\theta-\delta)}$ for some small $\delta > 0$. Applying the same proving strategy as for (35) to the truncated version of $\Psi_i(x)$, one can see that the arguments still go through. However, as the truncation points $\tau_{n,T}$ diverge much faster than in the standard case with fixed $i$, the convergence rate turns out to be slower than the standard rate $\sqrt{\log T/Th}$. $\Box$

**Lemma B2.** *Let $\Psi(x)$ be a kernel average of the form*

$$\Psi(x) = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} K_h(X_{it} - x) Z_{it}.$$

*It holds that*

$$\sup_{x\in[0,1]} \big|\Psi(x) - \mathbb{E}[\Psi(x)]\big| = O_p\Big(\sqrt{\frac{\log nT}{nTh}}\Big).$$

**Proof of Lemma B2.** As the proof closely follows standard arguments, we only provide a short sketch: Let $a_{n,T} = \sqrt{\log nT/nTh}$ and write $\Psi(x) = \Psi^{\leq}(x) + \Psi^{>}(x)$ with

$$\Psi^{\leq}(x) = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} K_h(X_{it}-x)Z_{it}I(|Z_{it}| \leq \tau_{n,T})$$
$$\Psi^{>}(x) = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} K_h(X_{it}-x)Z_{it}I(|Z_{it}| > \tau_{n,T}),$$

where the truncation sequence $\tau_{n,T}$ is given by $\tau_{n,T} = (nT)^{1/(\theta-\delta)}$ with some small $\delta > 0$. We thus have

$$\Psi(x) - \mathbb{E}[\Psi(x)] = (\Psi^{\leq}(x) - \mathbb{E}[\Psi^{\leq}(x)]) + (\Psi^{>}(x) - \mathbb{E}[\Psi^{>}(x)]).$$

Straightforward arguments show that $\sup_{x\in[0,1]} |\Psi^{>}(x) - \mathbb{E}[\Psi^{>}(x)]| = O_p(a_{n,T})$. To analyze the term $\sup_{x\in[0,1]} |\Psi^{\leq}(x) - \mathbb{E}[\Psi^{\leq}(x)]|$, we cover the unit interval by a grid of points $G_{n,T}$ that gets finer and finer as the sample size increases. We then replace the supremum over $x$ by the maximum over the grid points $x \in G_{n,T}$ and show that the resulting error is negligible. To complete the proof, we write

$$\mathbb{P}\Big(\max_{x\in G_{n,T}} \big|\Psi^{\leq}(x) - \mathbb{E}[\Psi^{\leq}(x)]\big| > Ca_{n,T}\Big) \leq \sum_{x\in G_{n,T}} \mathbb{P}\big(\big|\Psi^{\leq}(x) - \mathbb{E}[\Psi^{\leq}(x)]\big| > Ca_{n,T}\big)$$

and bound the probabilities $\mathbb{P}(|\Psi^{\leq}(x) - \mathbb{E}[\Psi^{\leq}(x)]| > Ca_{n,T})$ for each grid point with the help of an exponential inequality. To do so, let

$$\Psi^{\leq}(x) - \mathbb{E}[\Psi^{\leq}(x)] = \sum_{i=1}^{n}\sum_{t=1}^{T} W_{it}(x)$$

with $W_{it}(x) = \frac{1}{nT}\{K_h(X_{it}-x)Z_{it}I(|Z_{it}| \leq \tau_{n,T}) - \mathbb{E}[K_h(X_{it}-x)Z_{it}I(|Z_{it}| \leq \tau_{n,T})]\}$ and split up the expression $\sum_{t=1}^{T} W_{it}(x)$ into a growing number of blocks of increasing size. Using Bradley's lemma (see Lemma 1.2 in Bosq (1998)), we can replace these blocks by independent versions and apply an exponential inequality. $\square$

**Lemma B3.** *Let*

$$\Psi(x) = \frac{1}{n}\sum_{i=1}^{n} V_i(x)W_i(x)$$

*with*

$$V_i(x) = \Big(\frac{1}{T}\sum_{t=1}^{T}\big(K_h(X_{it}-x) - \mathbb{E}[K_h(X_{it}-x)]\big)\Big)^{\nu}$$
$$W_i(x) = \frac{1}{T}\sum_{t=1}^{T} K_h(X_{it}-x)Z_{it}$$

*for some fixed natural number $\nu$ and assume that the variables $Z_{it}$ satisfy $\mathbb{E}[Z_{it}|X_{it}] =$*

0. *Then*

$$\sup_{x\in[0,1]} \big|\Psi(x)\big| = o_p\Big(\frac{1}{\sqrt{nTh}}\Big).$$

**Proof of Lemma B3.** Throughout the proof, we use the following notation. Let

$\mathcal{C}_T$ : the event that $\max_i \sup_x |V_i(x)^{1/\nu}| \le C\sqrt{\log T/Th}$ and $\max_i \sup_x T^{-1}\sum_{t=1}^T K_h(X_{it}-x) \le C$

$\mathcal{C}_{iT}$ : the event that $\sup_x |V_i(x)^{1/\nu}| \le C\sqrt{\log T/Th}$ and $\sup_x T^{-1}\sum_{t=1}^T K_h(X_{it}-x) \le C$

for a fixed large constant $C$. Moreover, write $\mathcal{C}_T^c$ and $\mathcal{C}_{iT}^c$ to denote the complements of $\mathcal{C}_T$ and $\mathcal{C}_{iT}$, respectively. Inspecting the proof of Lemma B1, it is easily seen that $P(\mathcal{C}_T^c) = o(1)$ and $P(\mathcal{C}_{iT}^c) = o(1)$, given that the constant $C$ in the definition of the events $\mathcal{C}_T$ and $\mathcal{C}_{iT}$ is chosen sufficiently large. With this notation at hand, we obtain that

$$\begin{aligned}
\mathbb{P}\Big(\sup_{x\in[0,1]} \big|\Psi(x)\big| > Ma_{n,T}\Big) &\le \mathbb{P}\Big(\sup_{x\in[0,1]} \big|\Psi(x)\big| > Ma_{n,T}, \mathcal{C}_T\Big) \\
&\quad + \mathbb{P}\Big(\sup_{x\in[0,1]} \big|\Psi(x)\big| > Ma_{n,T}, \mathcal{C}_T^c\Big) \\
&= \mathbb{P}\Big(\sup_{x\in[0,1]} \big|\Psi(x)\big| > Ma_{n,T}, \mathcal{C}_T\Big) + o(1),
\end{aligned}$$

where $a_{n,T} = (\log nT\sqrt{nTh})^{-1}$ and $M$ is a large positive constant. Moreover,

$$\begin{aligned}
\mathbb{P}\Big(\sup_{x\in[0,1]} \big|\Psi(x)\big| > Ma_{n,T}, \mathcal{C}_T\Big) &= \mathbb{P}\Big(\sup_{x\in[0,1]} \Big|\frac{1}{n}\sum_{i=1}^n V_i(x)W_i(x)\Big| > Ma_{n,T}, \mathcal{C}_T\Big) \\
&= \mathbb{P}\Big(\sup_{x\in[0,1]} \Big|\frac{1}{n}\sum_{i=1}^n I(\mathcal{C}_T)V_i(x)W_i(x)\Big| > Ma_{n,T}\Big) \\
&\le \mathbb{P}\Big(\sup_{x\in[0,1]} \Big|\frac{1}{n}\sum_{i=1}^n I(\mathcal{C}_{iT})V_i(x)W_i(x)\Big| > Ma_{n,T}\Big).
\end{aligned}$$

Now write

$$\frac{1}{n}\sum_{i=1}^n I(\mathcal{C}_{iT})V_i(x)W_i(x) = Q^{\le}(x) + Q^{>}(x)$$

with the two terms on the right-hand side being defined as

$$\begin{aligned}
Q^{\le}(x) &= \frac{1}{n}\sum_{i=1}^n I(\mathcal{C}_{iT})V_i(x)W_i^{\le}(x) \\
Q^{>}(x) &= \frac{1}{n}\sum_{i=1}^n I(\mathcal{C}_{iT})V_i(x)W_i^{>}(x).
\end{aligned}$$

Here, $W_i(x) = W_i^{\leq}(x) + W_i^{>}(x)$ with

$$W_i^{\leq}(x) = \frac{1}{T}\sum_{t=1}^{T} K_h(X_{it}-x)Z_{it}^{\leq}$$
$$W_i^{>}(x) = \frac{1}{T}\sum_{t=1}^{T} K_h(X_{it}-x)Z_{it}^{>}$$

and $Z_{it} = Z_{it}^{\leq} + Z_{it}^{>}$ with

$$Z_{it}^{\leq} = Z_{it}I(|Z_{it}| \leq \tau_{n,T}) - \mathbb{E}[Z_{it}I(|Z_{it}| \leq \tau_{n,T})|X_{it}]$$
$$Z_{it}^{>} = Z_{it}I(|Z_{it}| > \tau_{n,T}) - \mathbb{E}[Z_{it}I(|Z_{it}| > \tau_{n,T})|X_{it}],$$

where the truncation sequence $\tau_{n,T}$ is chosen to equal $\tau_{n,T} = (nT)^{1/(\theta-\delta)}$ for some small $\delta > 0$. We now arrive at

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{x\in[0,1]}\Big|\frac{1}{n}\sum_{i=1}^{n} I(\mathcal{C}_{iT})V_i(x)W_i(x)\Big| > Ma_{n,T}\Big) \\
&\qquad \leq \mathbb{P}\Big(\sup_{x\in[0,1]}|Q^{\leq}(x)| > \frac{M}{2}a_{n,T}\Big) + \mathbb{P}\Big(\sup_{x\in[0,1]}|Q^{>}(x)| > \frac{M}{2}a_{n,T}\Big).
\end{aligned}$$

In the remainder of the proof, we show that the two terms on the right-hand side converge to zero as the sample size goes to infinity. To do so, we proceed in several steps.

*Step 1.* We start by considering the term $Q^{>}(x)$. It holds that

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{x\in[0,1]}\Big|\frac{1}{n}\sum_{i=1}^{n} I(\mathcal{C}_{iT})V_i(x)\Big(\frac{1}{T}\sum_{t=1}^{T} K_h(X_{it}-x)Z_{it}I(|Z_{it}| > \tau_{n,T})\Big)\Big| > Ca_{n,T}\Big) \\
&\leq \mathbb{P}\Big(|Z_{it}| > \tau_{n,T} \text{ for some } 1 \leq i \leq n \text{ and } 1 \leq t \leq T\Big) \\
&\leq \sum_{i=1}^{n}\sum_{t=1}^{T}\mathbb{P}(|Z_{it}| > \tau_{n,T}) \leq \sum_{i=1}^{n}\sum_{t=1}^{T}\mathbb{E}\Big[\frac{|Z_{it}|^{\theta}}{\tau_{n,T}^{\theta}}\Big] \leq C\frac{nT}{\tau_{n,T}^{\theta}} \to 0.
\end{aligned}$$

In addition,

$$\begin{aligned}
&\sup_{x\in[0,1]}\Big|\frac{1}{n}\sum_{i=1}^{n} I(\mathcal{C}_{iT})V_i(x)\Big(\frac{1}{T}\sum_{t=1}^{T} K_h(X_{it}-x)\mathbb{E}[Z_{it}I(|Z_{it}| > \tau_{n,T})|X_{it}]\Big)\Big| \\
&\leq C\sqrt{\frac{\log T}{Th}}\max_{1\leq i\leq n}\max_{1\leq t\leq T}\mathbb{E}\big[|Z_{it}|I(|Z_{it}| > \tau_{n,T})|X_{it}\big] \\
&\leq C\sqrt{\frac{\log T}{Th}}\frac{1}{\tau_{n,T}^{\theta-1}} \leq Ca_{n,T},
\end{aligned}$$

where the third line follows by (A4'). As a result,

$$\mathbb{P}\Big(\sup_{x\in[0,1]}|Q^{>}(x)| > \frac{M}{2}a_{n,T}\Big) = o(1)$$

for $M$ sufficiently large.

*Step 2.* We now turn to the analysis of the term $Q^{\leq}(x)$. Cover the region $[0,1]$ with open intervals $J_l$ ($l = 1,\ldots,L_{n,T}$) of length $C/L_{n,T}$ and let $x_l$ be the midpoint of the interval $J_l$. Then

$$\sup_{x\in[0,1]}|Q^{\leq}(x)| \leq \max_{1\leq l\leq L_{n,T}}|Q^{\leq}(x_l)| + \max_{1\leq l\leq L_{n,T}}\sup_{x\in J_l}|Q^{\leq}(x) - Q^{\leq}(x_l)|.$$

For any point $x \in J_l$, we have

$$I(\mathcal{C}_{iT})\big|V_i(x)W_i^{\leq}(x) - V_i(x_l)W_i^{\leq}(x_l)\big| \leq \frac{C\tau_{n,T}}{h^2}|x - x_l| \leq \frac{C\tau_{n,T}}{h^2 L_{n,T}}.$$

Therefore,

$$\max_{1\leq l\leq L_{n,T}}\sup_{x\in J_l}|Q^{\leq}(x) - Q^{\leq}(x_l)| \leq \frac{C\tau_{n,T}}{h^2 L_{n,T}}.$$

Choosing $L_{n,T} \to \infty$ with $L_{n,T} = C\tau_{n,T}/a_{n,T}h^2$, we obtain that

$$\max_{1\leq l\leq L_{n,T}}\sup_{x\in J_l}|Q^{\leq}(x) - Q^{\leq}(x_l)| \leq C a_{n,T}.$$

If we pick the constant $M$ large enough, we thus arrive at

$$\mathbb{P}\Big(\sup_{x\in[0,1]}|Q^{\leq}(x)| > \frac{M}{2}a_{n,T}\Big) \leq \mathbb{P}\Big(\max_{1\leq l\leq L_{n,T}}|Q^{\leq}(x_l)| > \frac{M}{4}a_{n,T}\Big) + o(1).$$

*Step 3.* It remains to show that

$$\mathbb{P}\Big(\max_{1\leq l\leq L_{n,T}}|Q^{\leq}(x_l)| > \frac{M}{4}a_{n,T}\Big) = o(1)$$

for some large fixed constant $M$. To do so, we write

$$\mathbb{P}\Big(\max_{1\leq l\leq L_{n,T}}|Q^{\leq}(x_l)| > \frac{M}{4}a_{n,T}\Big) \leq P_1 + P_2$$

with

$$P_1 = \mathbb{P}\Big(\max_{1\leq l\leq L_{n,T}}|Q^{\leq}(x_l) - \mathbb{E}Q^{\leq}(x_l)| > \frac{M}{8}a_{n,T}\Big)$$
$$P_2 = \mathbb{P}\Big(\max_{1\leq l\leq L_{n,T}}|\mathbb{E}Q^{\leq}(x_l)| > \frac{M}{8}a_{n,T}\Big).$$

First consider the term $P_2$. If $\nu \ge 3$, then

$$\begin{aligned}
|\mathbb{E}Q^{\le}(x_l)| &= \Big| \frac{1}{n} \sum_{i=1}^n \mathbb{E}[I(\mathcal{C}_{iT}) V_i(x_l) W_i^{\le}(x_l)] \Big| \\
&\le \frac{1}{n} \sum_{i=1}^n \mathbb{E}\big[I(\mathcal{C}_{iT}) V_i(x_l)^2\big]^{1/2} \mathbb{E}\big[W_i^{\le}(x_l)^2\big]^{1/2} \\
&\le \frac{C}{\sqrt{Th}} \Big(\frac{\log T}{Th}\Big)^{\nu/2} = o(a_{n,T}).
\end{aligned}$$

For $\nu \le 2$, we write

$$\begin{aligned}
|\mathbb{E}Q^{\le}(x_l)| &= \Big| \frac{1}{n} \sum_{i=1}^n \mathbb{E}[I(\mathcal{C}_{iT}) V_i(x_l) W_i^{\le}(x_l)] \Big| \\
&\le \Big| \frac{1}{n} \sum_{i=1}^n \mathbb{E}[V_i(x_l) W_i^{\le}(x_l)] \Big| + \Big| \frac{1}{n} \sum_{i=1}^n \mathbb{E}[I(\mathcal{C}_{iT}^c) V_i(x_l) W_i^{\le}(x_l)] \Big|.
\end{aligned}$$

If $\nu = 1$, we have

$$\begin{aligned}
\big|\mathbb{E}[V_i(x_l) W_i^{\le}(x_l)]\big| &= \Big| \frac{1}{T^2} \sum_{s,t=1}^T \mathbb{E}\big[(K_h(X_{is}-x_l) - \mathbb{E}[K_h(X_{is}-x_l)]) K_h(X_{it}-x_l) Z_{it}^{\le}\big] \Big| \\
&= \Big| \frac{1}{T^2} \sum_{\substack{s,t=1 \\ s \ne t}}^T \mathbb{E}\big[(K_h(X_{is}-x_l) - \mathbb{E}[K_h(X_{is}-x_l)]) K_h(X_{it}-x_l) Z_{it}^{\le}\big] \Big| \\
&\le \frac{C \log T}{T} = o(a_{n,T}),
\end{aligned}$$

the last line following with the help of Davydov's inequality and (A4'). For $\nu = 2$, it holds that

$$\begin{aligned}
\big|\mathbb{E}[V_i(x_l) W_i^{\le}(x_l)]\big| &= \Big| \frac{1}{T^3} \sum_{s,s',t=1}^T \mathbb{E}\big[(K_h(X_{is}-x_l) - \mathbb{E}[K_h(X_{is}-x_l)]) \\
&\qquad \times (K_h(X_{is'}-x_l) - \mathbb{E}[K_h(X_{is'}-x_l)]) K_h(X_{it}-x_l) Z_{it}^{\le}\big] \Big| \\
&\le \frac{CT(\log T)^2}{T^3 h^2} = C \Big(\frac{\log T}{Th}\Big)^2 = o(a_{n,T}),
\end{aligned}$$

the last line again following by Davydov's inequality and (A4'). In addition,

$$\mathbb{E}[I(\mathcal{C}_{iT}^c) V_i(x_l) W_i^{\le}(x_l)] \le \mathbb{E}[I(\mathcal{C}_{iT}^c)]^{1/2} \mathbb{E}[V_i(x_l)^2 W_i^{\le}(x_l)^2]^{1/2}.$$

Repeating the usual strategy to prove uniform convergence for kernel estimates, it can be shown that under our assumptions, $\mathbb{E}[I(\mathcal{C}_{iT}^c)] = \mathbb{P}(\mathcal{C}_{iT}^c) \le T^{-C}$ for an arbitrarily large constant $C$. This yields that $\mathbb{E}[I(\mathcal{C}_{iT}^c) V_i(x_l) W_i^{\le}(x_l)] = o(a_{n,T})$, which in turn implies that $|\mathbb{E}Q^{\le}(x_l)| = o(a_{n,T})$ for $\nu = 1, 2$. As a result, $P_2 = o(1)$ for any $\nu \ge 1$.

To cope with the term $P_1$, we apply the bound

$$P_1 \le \sum_{l=1}^{L_{n,T}} \mathbb{P}\Big( |Q^{\le}(x_l) - \mathbb{E}Q^{\le}(x_l)| > \frac{M}{8} a_{n,T} \Big)$$

and consider the probability $\mathbb{P}(|Q^{\le}(x_l) - \mathbb{E}Q^{\le}(x_l)| > M a_{n,T}/8)$ for an arbitrary fixed grid point $x_l$. Write

$$Q^{\le}(x_l) - \mathbb{E}Q^{\le}(x_l) = \sum_{i=1}^{n} \xi_i(x_l)$$

with $\xi_i(x_l) = n^{-1}\{I(\mathcal{C}_{iT})V_i(x_l)W_i^{\le}(x_l) - \mathbb{E}[I(\mathcal{C}_{iT})V_i(x_l)W_i^{\le}(x_l)]\}$. Recalling the definition of the events $\mathcal{C}_{iT}$, the variables $\xi_i(x_l)$ can be bounded as follows:

$$|\xi_i(x_l)| \le C \sqrt{\frac{\log T}{Th}} \frac{\tau_{n,T}}{n} \le \frac{C}{(nTh)^{1/2+\delta}} := \overline{C}_{n,T}$$

with some sufficiently large constant $C$ and a small $\delta > 0$, given that $n \gg T^{2/3}$ and $\theta > 5$. With $\lambda_{n,T} = \overline{C}_{n,T}^{-1}/2$, we obtain that $\lambda_{n,T}|\xi_i(x_l)| \le 1/2$. As $\exp(x) \le 1+x+x^2$ for $|x| \le 1/2$,

$$\mathbb{E}\Big[ \exp\big( \pm \lambda_{n,T} \xi_i(x_l) \big) \Big] \le 1 + \lambda_{n,T}^2 \mathbb{E}[\xi_i(x_l)^2] \le \exp\Big( \lambda_{n,T}^2 \mathbb{E}[\xi_i(x_l)^2] \Big).$$

Using this together with Markov's inequality, we arrive at

$$\begin{aligned}
&\mathbb{P}\Big( \Big| \sum_{i=1}^{n} \xi_i(x_l) \Big| > \frac{M}{8} a_{n,T} \Big) \\
&\le \exp\Big( -\frac{M}{8} \lambda_{n,T} a_{n,T} \Big) \Big\{ \mathbb{E}\Big[ \exp\big( \lambda_{n,T} \sum_{i=1}^{n} \xi_i(x_l) \big) \Big] + \mathbb{E}\Big[ \exp\big( -\lambda_{n,T} \sum_{i=1}^{n} \xi_i(x_l) \big) \Big] \Big\} \\
&\le 2 \exp\Big( -\frac{M}{8} \lambda_{n,T} a_{n,T} \Big) \prod_{i=1}^{n} \exp\big( \lambda_{n,T}^2 \mathbb{E}[\xi_i(x_l)^2] \big) \\
&= 2 \exp\Big( -\frac{M}{8} \lambda_{n,T} a_{n,T} \Big) \exp\Big( \lambda_{n,T}^2 \sum_{i=1}^{n} \mathbb{E}[\xi_i(x_l)^2] \Big).
\end{aligned}$$

Now note that

$$\mathbb{E}[\xi_i(x_l)^2] \le \frac{1}{n^2} \mathbb{E}[I(\mathcal{C}_{iT}) V_i(x_l)^2 W_i^{\le}(x_l)^2] \le \frac{C \log T}{n^2 Th} \mathbb{E}[W_i^{\le}(x_l)^2]$$

and

$$\mathbb{E}[W_i^{\le}(x_l)^2] = \frac{1}{T^2} \sum_{s,t=1}^{T} \mathbb{E}\big[ K_h(X_{is} - x_l) K_h(X_{it} - x_l) Z_{is}^{\le} Z_{it}^{\le} \big]$$

$$= \frac{1}{T^2} \sum_{s,t=1}^{T} \mathrm{Cov}\big(K_h(X_{is} - x_l) Z_{is}^{\leq}, K_h(X_{it} - x_l) Z_{it}^{\leq}\big) \leq \frac{C}{Th}.$$

Hence, $\mathbb{E}[\xi_i(x_l)^2] \leq C \log T/(nTh)^2$ and

$$\lambda_{n,T}^2 \sum_{i=1}^{n} \mathbb{E}[\xi_i(x_l)^2] \leq C(nTh)^{1+2\delta} \frac{\log T}{n(Th)^2} \leq C \frac{(nT)^{2\delta}}{Th} = o(1).$$

Moreover,

$$\lambda_{n,T} a_{n,T} = \frac{(nTh)^{1/2+\delta}}{\log nT (nTh)^{1/2}} \to \infty$$

at polynomial rate. As a result,

$$\mathbb{P}\Big( \Big| \sum_{i=1}^{n} \xi_i(x_l) \Big| > \frac{M}{8} a_{n,T} \Big) \leq C T^{-p},$$

where the constant $p > 0$ can be chosen arbitrarily large. This completes the proof. $\square$

**Lemma B4.** *Let*

$$\Psi(x) = \frac{1}{T} \sum_{t=1}^{T} V_t(x) W_t,$$

*where* $W_t = \frac{1}{n} \sum_{i=1}^{n} Z_{it}$ *and*

$$V_t(x) = \frac{1}{n} \sum_{i=1}^{n} \big( K_h(X_{it} - x) - \mathbb{E}[K_h(X_{it} - x)] \big).$$

*Assume that the variables* $Z_{it}$ *have mean zero. Then it holds that*

$$\sup_{x \in [0,1]} \big| \Psi(x) \big| = o_p\Big( \frac{1}{\sqrt{nTh}} \Big).$$

**Proof of Lemma B4.** The proof is similar to that of Lemma B3 with the roles of $i$ and $t$ being reversed. Let $a_{n,T} = (\log nT \sqrt{nTh})^{-1}$ and $\tau_{n,T} = (nT)^{1/(\theta - \delta)}$ for some small $\delta > 0$. Arguments analogous to those for Step 1 in the proof of Lemma B3 yield that $\Psi(x)$ can be replaced by the term

$$Q^{\leq}(x) = \frac{1}{T} \sum_{t=1}^{T} I(\mathcal{C}_{tn}) V_t(x) W_t^{\leq},$$

where $W_t^{\leq} = \frac{1}{n} \sum_{i=1}^{n} Z_{it}^{\leq}$ with $Z_{it}^{\leq} = Z_{it} I(|Z_{it}| \leq \tau_{n,T}) - \mathbb{E}[Z_{it} I(|Z_{it}| \leq \tau_{n,T})]$ and $\mathcal{C}_{tn}$ is the event that $\sup_x |V_t(x)| \leq C \sqrt{\log n/nh}$ for some sufficiently large constant $C$. Next cover the unit interval by a grid of $L_{n,T} = C \tau_{n,T}/a_{n,T} h^2$ points. As in the proof of Lemma B3, we can show that

$$\sup_{x\in[0,1]} |Q^{\leq}(x)| = \max_{1\leq l\leq L_{n,T}} |Q^{\leq}(x_l)| + O(a_{n,T}).$$

Moreover, again repeating the arguments from Lemma B3, we obtain that for some sufficiently large constant $M$,

$$\begin{aligned}\mathbb{P}\Big(\max_{1\leq l\leq L_{n,T}} |Q^{\leq}(x_l)| > M a_{n,T}\Big) &\leq \mathbb{P}\Big(\max_{1\leq l\leq L_{n,T}} |Q^{\leq}(x_l) - \mathbb{E}Q^{\leq}(x_l)| > \frac{M}{2} a_{n,T}\Big) + o(1) \\ &\leq \sum_{l=1}^{L_{n,T}} \mathbb{P}\Big(|Q^{\leq}(x_l) - \mathbb{E}Q^{\leq}(x_l)| > \frac{M}{2} a_{n,T}\Big) + o(1).\end{aligned}$$

To complete the proof, we bound the probability $\mathbb{P}(|Q^{\leq}(x) - \mathbb{E}Q^{\leq}(x)| > \frac{M}{2} a_{n,T})$ for an arbitrary point $x$ by an exponential inequality. To do so, we must slightly vary the arguments for Lemma B3, taking into account the fact that $Q^{\leq}(x)$ is not a sum of independent terms any more. In particular, we write

$$Q^{\leq}(x) - \mathbb{E}Q^{\leq}(x) = \sum_{t=1}^{T} \xi_t(x)$$

with $\xi_t(x) = T^{-1}\{I(\mathcal{C}_{tn})V_t(x)W_t^{\leq} - \mathbb{E}[I(\mathcal{C}_{tn})V_t(x)W_t^{\leq}]\}$ and split up the expression $\sum_{t=1}^{T} \xi_t(x)$ into blocks as follows:

$$\sum_{t=1}^{T} \xi_t(x) = \sum_{s=1}^{q_{n,T}} B_{2s-1}(x) + \sum_{s=1}^{q_{n,T}} B_{2s}(x)$$

with $B_s(x) = \sum_{t=(s-1)r_{n,T}+1}^{s r_{n,T}} \xi_t(x)$, where $2q_{n,T}$ is the number of blocks and $r_{n,T} = T/2q_{n,T}$ is the block length. We now get

$$\begin{aligned}\mathbb{P}\Big(\Big|\sum_{t=1}^{T} \xi_t(x)\Big| > \frac{M}{2} a_{n,T}\Big) &\leq \mathbb{P}\Big(\Big|\sum_{s=1}^{q_{n,T}} B_{2s-1}(x)\Big| > \frac{M}{4} a_{n,T}\Big) \\ &\quad + \mathbb{P}\Big(\Big|\sum_{s=1}^{q_{n,T}} B_{2s}(x)\Big| > \frac{M}{4} a_{n,T}\Big).\end{aligned}$$

In what follows, we restrict attention to the first term on the right-hand side of the above display. The second one can be analyzed by analogous arguments. We make use of the following two facts:

(1) Let $\mathcal{V}^{(i)} = \{\mathcal{V}_t^{(i)} : t = 1, \ldots, T\} = \{(X_{it}, Z_{it}) : t = 1, \ldots, T\}$ be the time series of the $i$-th individual and consider the time series $\mathcal{W} = \{\mathcal{W}_t : t = 1, \ldots, T\}$ with $\mathcal{W}_t = h_t(\mathcal{V}_t^{(1)}, \ldots, \mathcal{V}_t^{(n)}) = h_t(X_{1t}, Z_{1t}, \ldots, X_{nt}, Z_{nt})$ for some Borel functions $h_t$. Then by Theorem 5.2 in Bradley (2005) and the comments thereafter, the mixing coefficients $\alpha^{\mathcal{W}}(k)$ of the time series $\mathcal{W}$ are such that $\alpha^{\mathcal{W}}(k) \leq \sum_{i=1}^{n} \alpha_i(k) \leq$

$n\alpha(k)$ for each $k \in \mathbb{N}$. In particular, letting $\alpha^{\xi}(k)$ be the mixing coefficients of the time series $\{\xi_t(x)\}$, it holds that $\alpha^{\xi}(k) \le n\alpha(k)$.

(2) By Bradley's lemma (see Lemma 1.2 in Bosq (1998), we can construct a sequence of random variables $B_1^*(x), B_3^*(x), \ldots$ such that (i) $B_1^*(x)$, $B_3^*(x)$, $\ldots$ are independent, (ii) $B_{2s-1}^*(x)$ has the same distribution as $B_{2s-1}(x)$, and (iii) for $0 < \mu \le \|B_{2s-1}(x)\|_\infty$, it holds that

$$\mathbb{P}\big(|B_{2s-1}^*(x) - B_{2s-1}(x)| > \mu\big) \le 18\Big(\frac{\|B_{2s-1}(x)\|_\infty}{\mu}\Big)^{1/2} \alpha^{\xi}(r_{n,T}). \qquad \text{(A.15)}$$

Using fact (2), we can write

$$\mathbb{P}\Big(\Big|\sum_{s=1}^{q_{n,T}} B_{2s-1}(x)\Big| > \frac{M}{4} a_{n,T}\Big) \le P_1 + P_2$$

with

$$\begin{aligned} P_1 &= \mathbb{P}\Big(\Big|\sum_{s=1}^{q_{n,T}} B_{2s-1}^*(x)\Big| > \frac{M}{8} a_{n,T}\Big) \\ P_2 &= \mathbb{P}\Big(\Big|\sum_{s=1}^{q_{n,T}} \big(B_{2s-1}(x) - B_{2s-1}^*(x)\big)\Big| > \frac{M}{8} a_{n,T}\Big). \end{aligned}$$

We first consider $P_1$. Picking the block length to equal $r_{n,T} = (nT)^{\eta}$ for some small $\eta > 0$, it holds that $|B_{2s-1}(x)| \le C\sqrt{\frac{\log n}{nh}\frac{\tau_{n,T} r_{n,T}}{T}} \le \frac{C}{(nTh)^{1/2+\delta}} =: \overline{C}_{n,T}$ with some sufficiently large constant $C$ and a small $\delta > 0$. Choosing $\lambda_{n,T} = \overline{C}_{n,T}^{-1}/2$ and applying Markov's inequality, the same arguments as in Lemma B3 yield that

$$P_1 \le 2\exp\Big(-\frac{M}{8}\lambda_{n,T} a_{n,T} + \lambda_{n,T}^2 \sum_{s=1}^{q_{n,T}} \mathbb{E}[B_{2s-1}^*(x)^2]\Big).$$

Since $\sum_{s=1}^{q_{n,T}} \mathbb{E}[B_{2s-1}^*(x)^2] \le C \log n \log T / n^2 Th$, we finally arrive at

$$P_1 \le 2\exp\Big(-\frac{M}{8}\lambda_{n,T} a_{n,T} + C\lambda_{n,T}^2 \frac{\log n \log T}{n^2 Th}\Big).$$

Direct calculations show that $\lambda_{n,T} a_{n,T} \to \infty$, whereas $\lambda_{n,T}^2 \frac{\log n \log T}{n^2 Th} = o(1)$. This implies that $P_1$ converges to zero at an arbitarily fast polynomial rate. Moreover, using (A.15) together with the fact that $\alpha^{\xi}(k) \le n\alpha(k)$ and recalling that the coefficients $\alpha(k)$ decay exponentially fast to zero, it immediately follows that $P_2$ converges to zero at an arbitrarily fast polynomial rate as well. From this, the result easily follows. $\square$

**Lemma B5.** *Let*

$$\Psi(x) = \frac{1}{n}\sum_{i=1}^{n}\Big(\frac{1}{nT}\sum_{\substack{j=1\\ j\neq i}}^{n}\sum_{t=1}^{T}\varphi_{it}(x)Z_{jt}\Big)$$

*with* $\varphi_{it}(x) = K_h(X_{it}-x)\widehat{\phi}_i(x)$ *and* $\widehat{\phi}_i(x)$ *an estimator based on the data* $\{X_{it} : t = 1,\ldots,T\}$*. Assume that* $\widehat{\phi}_i(x)$ *has the following two properties:*

*(a)* $\mathbb{P}(\max_{1\le i\le n}\sup_{x\in[0,1]}|\widehat{\phi}_i(x)| > Cb_{n,T}) = o(1)$ *for a sufficiently large constant* $C$ *and a null sequence* $\{b_{n,T}\}$ *which satisfies* $b_{n,T}^2/h \le C(nT)^{-\eta}$ *for some small* $\eta > 0$.

*(b)* $\max_{1\le i\le n}|\widehat{\phi}_i(x)-\widehat{\phi}_i(x')| \le c_{n,T}|x-x'|$ *with probability tending to one for some sequence* $\{c_{n,T}\}$ *which satisfies* $c_{n,T} \le (nT)^C$ *for some positive constant* $C$.

*In addition, let the variables* $Z_{it}$ *have mean zero. Then it holds that*

$$\sup_{x\in[0,1]}\big|\Psi(x)\big| = o_p\Big(\frac{1}{\sqrt{nTh}}\Big).$$

**Proof of Lemma B5.** Let $\mathcal{C}_T$ be the event that $\max_{1\le i\le n}\sup_{x\in[0,1]}|\widehat{\phi}_i(x)| \le Cb_{n,T}$ and $\mathcal{C}_{iT}$ the event that $\sup_{x\in[0,1]}|\widehat{\phi}_i(x)| \le Cb_{n,T}$. Moreover, write $\mathcal{C}_T^c$ and $\mathcal{C}_{iT}^c$ to denote the complements of $\mathcal{C}_T$ and $\mathcal{C}_{iT}$, respectively. By assumption, $P(\mathcal{C}_T^c) = o(1)$ and $P(\mathcal{C}_{iT}^c) = o(1)$. With this notation at hand, we have

$$\begin{aligned}\mathbb{P}\Big(\sup_{x\in[0,1]}\big|\Psi(x)\big| > Ma_{n,T}\Big) &\le \mathbb{P}\Big(\sup_{x\in[0,1]}\big|\Psi(x)\big| > Ma_{n,T}, \mathcal{C}_T\Big)\\ &\quad + \mathbb{P}\Big(\sup_{x\in[0,1]}\big|\Psi(x)\big| > Ma_{n,T}, \mathcal{C}_T^c\Big)\\ &\le \mathbb{P}\Big(\sup_{x\in[0,1]}\big|\Psi(x)\big| > Ma_{n,T}, \mathcal{C}_T\Big) + o(1),\end{aligned}$$

where $a_{n,T} = \sqrt{\frac{\log nT}{nTh(nT)^\eta}}$ and $M$ is a positive constant. Moreover,

$$\begin{aligned}&\mathbb{P}\Big(\sup_{x\in[0,1]}\big|\Psi(x)\big| > Ma_{n,T}, \mathcal{C}_T\Big)\\ &= \mathbb{P}\Big(\sup_{x\in[0,1]}\Big|\frac{1}{n}\sum_{i=1}^{n}\Big(\frac{1}{nT}\sum_{j\neq i}\sum_{t=1}^{T}I(\mathcal{C}_T)\varphi_{it}(x)Z_{jt}\Big)\Big| > Ma_{n,T}\Big)\\ &\le \mathbb{P}\Big(\sup_{x\in[0,1]}\Big|\frac{1}{n}\sum_{i=1}^{n}\Big(\frac{1}{nT}\sum_{j\neq i}\sum_{t=1}^{T}I(\mathcal{C}_{iT})\varphi_{it}(x)Z_{jt}\Big)\Big| > Ma_{n,T}\Big).\end{aligned}$$

Defining

$$\begin{aligned}Z_{jt}^{\le} &= Z_{jt}I(|Z_{jt}| \le \tau_{n,T}) - \mathbb{E}\big[Z_{jt}I(|Z_{jt}| \le \tau_{n,T})\big]\\ Z_{jt}^{>} &= Z_{jt}I(|Z_{jt}| > \tau_{n,T}) - \mathbb{E}\big[Z_{jt}I(|Z_{jt}| > \tau_{n,T})\big]\end{aligned}$$

with $\tau_{n,T} = (nT)^{1/(\theta-\delta)}$ for some small $\delta > 0$, we further get that

$$\frac{1}{n}\sum_{i=1}^{n}\Big(\frac{1}{nT}\sum_{j\neq i}\sum_{t=1}^{T} I(\mathcal{C}_{iT})\varphi_{it}(x)Z_{jt}\Big) = Q^{\leq}(x) + Q^{>}(x)$$

with

$$\begin{aligned}
Q^{\leq}(x) &= \frac{1}{n}\sum_{i=1}^{n}\Big(\frac{1}{nT}\sum_{j\neq i}\sum_{t=1}^{T} I(\mathcal{C}_{iT})\varphi_{it}(x)Z_{jt}^{\leq}\Big) \\
Q^{>}(x) &= \frac{1}{n}\sum_{i=1}^{n}\Big(\frac{1}{nT}\sum_{j\neq i}\sum_{t=1}^{T} I(\mathcal{C}_{iT})\varphi_{it}(x)Z_{jt}^{>}\Big).
\end{aligned}$$

Hence,

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{x\in[0,1]}\Big|\frac{1}{n}\sum_{i=1}^{n}\Big(\frac{1}{nT}\sum_{j\neq i}\sum_{t=1}^{T} I(\mathcal{C}_{iT})\varphi_{it}(x)Z_{jt}\Big)\Big| > Ma_{n,T}\Big) \\
&\quad\leq \mathbb{P}\Big(\sup_{x\in[0,1]}\big|Q^{\leq}(x)\big| > \frac{M}{2}a_{n,T}\Big) + \mathbb{P}\Big(\sup_{x\in[0,1]}\big|Q^{>}(x)\big| > \frac{M}{2}a_{n,T}\Big).
\end{aligned}$$

In what follows, we show that the two terms on the right-hand side converge to zero as the sample size increases. The proof splits up into several steps.

*Step 1.* We first consider $Q^{>}(x)$. Similarly to Lemma B3, it holds that

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{x\in[0,1]}\Big|\frac{1}{n}\sum_{i=1}^{n}\Big(\frac{1}{nT}\sum_{j\neq i}\sum_{t=1}^{T} I(\mathcal{C}_{iT})\varphi_{it}(x)Z_{jt}I(|Z_{jt}| > \tau_{n,T})\Big)\Big| > Ca_{n,T}\Big) \\
&\leq \mathbb{P}\Big(|Z_{jt}| > \tau_{n,T} \text{ for some } 1 \leq j \leq n \text{ and } 1 \leq t \leq T\Big) \to 0
\end{aligned}$$

and

$$\begin{aligned}
&\sup_{x\in[0,1]}\Big|\frac{1}{n}\sum_{i=1}^{n}\Big(\frac{1}{nT}\sum_{j\neq i}\sum_{t=1}^{T} I(\mathcal{C}_{iT})\varphi_{it}(x)\mathbb{E}\big[Z_{jt}I(|Z_{jt}| > \tau_{n,T})\big]\Big)\Big| \\
&\leq \frac{Cb_{n,T}}{n^2Th}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{t=1}^{T}\mathbb{E}\big[|Z_{jt}|I(|Z_{jt}| > \tau_{n,T})\big] \leq \frac{Cb_{n,T}}{\tau_{n,T}^{\theta-1}h} \leq Ca_{n,T}.
\end{aligned}$$

From this, it immediately follows that $\mathbb{P}(\sup_{x\in[0,1]}|Q^{>}(x)| > Ma_{n,T}/2) = o(1)$ for $M$ sufficiently large.

*Step 2.* We now turn to the analysis of $Q^{\leq}(x)$. Let $L_{n,T} \to \infty$ with $L_{n,T} = \max\{\frac{\tau_{n,T}c_{n,T}}{ha_{n,T}}, \frac{b_{n,T}\tau_{n,T}}{h^2a_{n,T}}, (nT)^{\delta}\}$ for some small $\delta > 0$. Cover the region $[0,1]$ with open intervals $J_l$ ($l = 1, \dots, L_{n,T}$) of length $C/L_{n,T}$ and let $x_l$ be the midpoint of the interval $J_l$. Then for $x \in J_l$,

$$\begin{aligned}
\big|Q^{\leq}(x) - Q^{\leq}(x_l)\big| &\leq \frac{C\tau_{n,T}}{n}\sum_{i=1}^{n}\Big(\frac{1}{nT}\sum_{j\neq i}\sum_{t=1}^{T} I(\mathcal{C}_{iT})|\varphi_{it}(x) - \varphi_{it}(x_l)|\Big) \\
&\leq \frac{C\tau_{n,T}}{n}\sum_{i=1}^{n}\Big(\frac{1}{nT}\sum_{j\neq i}\sum_{t=1}^{T} I(\mathcal{C}_{iT})\big\{K_h(X_{it}-x)|\widehat{\phi}_i(x) - \widehat{\phi}_i(x_l)| \\
&\qquad\qquad + |\widehat{\phi}_i(x_l)||K_h(X_{it}-x) - K_h(X_{it}-x_l)|\big\}\Big) \\
&\leq C\tau_{n,T}\Big(\frac{c_{n,T}}{h} + \frac{b_{n,T}}{h^2}\Big)|x - x_l| \leq C\frac{\tau_{n,T}}{L_{n,T}}\Big(\frac{c_{n,T}}{h} + \frac{b_{n,T}}{h^2}\Big) \leq Ca_{n,T}
\end{aligned}$$

with probability tending to one. From this, it immediately follows that for sufficiently large $M$,

$$\mathbb{P}\Big(\sup_{x\in[0,1]}\big|Q^{\leq}(x)\big| > \frac{M}{2}a_{n,T}\Big) \leq \mathbb{P}\Big(\max_{1\leq l\leq L_{n,T}}\big|Q^{\leq}(x_l)\big| > \frac{M}{4}a_{n,T}\Big) + o(1).$$

*Step 3.* It remains to show that

$$\mathbb{P}\Big(\max_{1\leq l\leq L_{n,T}}\big|Q^{\leq}(x_l)\big| > \frac{M}{4}a_{n,T}\Big) = o(1)$$

for some sufficiently large constant $M$. Writing

$$\max_{1\leq l\leq L_{n,T}}\big|Q^{\leq}(x_l)\big| \leq \max_{\substack{1\leq i\leq n\\ 1\leq l\leq L_{n,T}}}\Big|\sum_{j\neq i}\sum_{t=1}^{T} I(\mathcal{C}_{iT})\varphi_{it}(x_l)W_{jt}\Big|$$

with $W_{jt} = \frac{1}{nT}\{Z_{jt}I(|Z_{jt}| \leq \tau_{n,T}) - \mathbb{E}[Z_{jt}I(|Z_{jt}| \leq \tau_{n,T})]\}$, we obtain

$$\begin{aligned}
&\mathbb{P}\Big(\max_{1\leq l\leq L_{n,T}}\big|Q^{\leq}(x_l)\big| > \frac{M}{4}a_{n,T}\Big) \\
&\leq \mathbb{P}\Big(\max_{\substack{1\leq i\leq n\\ 1\leq l\leq L_{n,T}}}\Big|\sum_{j\neq i}\sum_{t=1}^{T} I(\mathcal{C}_{iT})\varphi_{it}(x_l)W_{jt}\Big| > \frac{M}{4}a_{n,T}\Big) \\
&\leq \sum_{i=1}^{n}\sum_{l=1}^{L_{n,T}}\mathbb{P}\Big(\Big|\sum_{j\neq i}\sum_{t=1}^{T} I(\mathcal{C}_{iT})\varphi_{it}(x_l)W_{jt}\Big| > \frac{M}{4}a_{n,T}\Big).
\end{aligned}$$

We now bound the probability $\mathbb{P}(|\sum_{j\neq i}\sum_{t=1}^{T} I(\mathcal{C}_{iT})\varphi_{it}(x)W_{jt}| > Ma_{n,T}/4)$ for an arbitrary point $x$ with the help of an exponential inequality. To do so, we rewrite the expression $\sum_{j\neq i}\sum_{t=1}^{T} I(\mathcal{C}_{iT})\varphi_{it}(x)W_{jt}$. In particular, we split up the inner sum over $t$ into blocks as follows:

$$\sum_{t=1}^{T} I(\mathcal{C}_{iT})\varphi_{it}(x)W_{jt} = \sum_{s=1}^{q_{n,T}} B_{j,2s-1}(x) + \sum_{s=1}^{q_{n,T}} B_{j,2s}(x)$$

with

$$B_{j,s}(x) = \sum_{t=(s-1)r_{n,T}+1}^{sr_{n,T}} I(\mathcal{C}_{iT})\varphi_{it}(x)W_{jt},$$

where as in Lemma B4, $2q_{n,T}$ is the number of blocks and $r_{n,T} = T/2q_{n,T}$ is the block length. We thus get

$$\begin{aligned}\mathbb{P}\Big(\Big|\sum_{j\neq i}\sum_{t=1}^{T} I(\mathcal{C}_{iT})\varphi_{it}(x)W_{jt}\Big| > \frac{M}{4}a_{n,T}\Big) &\leq \mathbb{P}\Big(\Big|\sum_{j\neq i}\sum_{s=1}^{q_{n,T}} B_{j,2s-1}(x)\Big| > \frac{M}{8}a_{n,T}\Big) \\ &\quad + \mathbb{P}\Big(\Big|\sum_{j\neq i}\sum_{s=1}^{q_{n,T}} B_{j,2s}(x)\Big| > \frac{M}{8}a_{n,T}\Big).\end{aligned}$$

In what follows, we restrict attention to the first term on the right-hand side. The second one can be analyzed by similar arguments.

To indicate the dependence of the block $B_{j,s}(x)$ on the $i$-th time series $\{X_{it}\}_{t=1}^{T}$, we use the notation $B_{j,s}(x) = B_{j,s}(x, \{X_{it}\}_{t=1}^{T})$. Moreover, we employ the shorthand $\overline{B}_{j,s}(x) = B_{j,s}(x, \{x_{it}\}_{t=1}^{T})$ to denote the $s$-th block for a fixed realization $\{x_{it}\}_{t=1}^{T}$ of $\{X_{it}\}_{t=1}^{T}$. With this notation at hand, we write

$$\begin{aligned}&\mathbb{P}\Big(\Big|\sum_{j\neq i}\sum_{s=1}^{q_{n,T}} B_{j,2s-1}(x)\Big| > \frac{M}{8}a_{n,T}\Big) \\ &\qquad = \mathbb{E}\Big[\mathbb{P}\Big(\Big|\sum_{j\neq i}\sum_{s=1}^{q_{n,T}} B_{j,2s-1}(x)\Big| > \frac{M}{8}a_{n,T}\Big|\{X_{it}\}_{t=1}^{T}\Big)\Big]\end{aligned}$$

and bound the term

$$\begin{aligned}&\mathbb{P}\Big(\Big|\sum_{j\neq i}\sum_{s=1}^{q_{n,T}} B_{j,2s-1}(x)\Big| > \frac{M}{8}a_{n,T}\Big|\{X_{it}\}_{t=1}^{T} = \{x_{it}\}_{t=1}^{T}\Big) \\ &\qquad\qquad = \mathbb{P}\Big(\Big|\sum_{j\neq i}\sum_{s=1}^{q_{n,T}} \overline{B}_{j,2s-1}(x)\Big| > \frac{M}{8}a_{n,T}\Big)\end{aligned}$$

for an arbitrary but fixed realization $\{x_{it}\}_{t=1}^{T}$. By Bradley's lemma, we can construct a sequence of random variables $\overline{B}^*_{j,1}(x), \overline{B}^*_{j,3}(x), \ldots$ such that (i) $\overline{B}^*_{j,1}(x), \overline{B}^*_{j,3}(x), \ldots$ are independent, (ii) $\overline{B}^*_{j,2s-1}(x)$ has the same distribution as $\overline{B}_{j,2s-1}(x)$, and (iii) for $0 < \mu \leq \|\overline{B}_{j,2s-1}(x)\|_\infty$,

$$\mathbb{P}\big(|\overline{B}^*_{j,2s-1}(x) - \overline{B}_{j,2s-1}(x)| > \mu\big) \leq 18\Big(\frac{\|\overline{B}_{j,2s-1}(x)\|_\infty}{\mu}\Big)^{1/2}\alpha(r_{n,T}). \tag{A.16}$$

This allows us to write

$$\mathbb{P}\Big(\Big|\sum_{j\neq i}\sum_{s=1}^{q_{n,T}} \overline{B}_{j,2s-1}(x)\Big| > \frac{M}{8}a_{n,T}\Big) \leq P_1 + P_2$$

with

$$P_1 = \mathbb{P}\Big( \Big| \sum_{j \neq i} \sum_{s=1}^{q_{n,T}} \overline{B}^*_{j,2s-1}(x) \Big| > \frac{M}{16} a_{n,T} \Big)$$

$$P_2 = \mathbb{P}\Big( \Big| \sum_{j \neq i} \sum_{s=1}^{q_{n,T}} \big( \overline{B}_{j,2s-1}(x) - \overline{B}^*_{j,2s-1}(x) \big) \Big| > \frac{M}{16} a_{n,T} \Big).$$

First consider $P_1$. It holds that

$$|\overline{B}_{j,2s-1}(x)| \leq \frac{C\tau_{n,T} r_{n,T} b_{n,T}}{nTh} \leq \frac{C\tau_{n,T} r_{n,T} (b_{n,T}/\sqrt{h})}{nTh} \leq \frac{C\tau_{n,T} r_{n,T}}{nTh(nT)^{\eta/2}} =: \overline{C}_{n,T}.$$

Choosing $\lambda_{n,T} = \overline{C}_{n,T}^{-1}/2$ and applying Markov's inequality, the same arguments as in Lemma B3 yield that

$$P_1 \leq 2 \exp\Big( - \frac{M}{16} \lambda_{n,T} a_{n,T} + \lambda_{n,T}^2 \sum_{j \neq i} \sum_{s=1}^{q_{n,T}} \mathbb{E}[\overline{B}^*_{j,2s-1}(x)^2] \Big).$$

Noting that

$$\begin{aligned}
\sum_{j \neq i} \sum_{s=1}^{q_{n,T}} \mathbb{E}[\overline{B}^*_{j,2s-1}(x)^2] &= \sum_{j \neq i} \sum_{s=1}^{q_{n,T}} \mathbb{E}[\overline{B}_{j,2s-1}(x)^2] \\
&= \sum_{j \neq i} \sum_{s=1}^{q_{n,T}} \mathbb{E}[B_{j,2s-1}(x)^2 | \{X_{it}\}_{t=1}^T = \{x_{it}\}_{t=1}^T] \\
&\leq \sum_{j \neq i} \sum_{s,t=1}^{T} I(\mathcal{C}_{iT}) |\varphi_{is}(x)\varphi_{it}(x)| \big| \mathbb{E}[W_{js} W_{jt}] \big| \\
&\leq C b_{n,T}^2 \sum_{j \neq i} \sum_{s,t=1}^{T} K_h(x_{is} - x) K_h(x_{it} - x) \big| \mathbb{E}[W_{js} W_{jt}] \big| \\
&\leq \frac{C b_{n,T}^2}{h^2} \sum_{j \neq i} \Big( \sum_{t=1}^{T} |\mathbb{E}[W_{jt}^2]| + 2 \sum_{l=1}^{T-1} \sum_{t=1}^{T-l} |\mathbb{E}[W_{jt} W_{jt+l}]| \Big) \\
&\leq \frac{C}{nTh(nT)^\eta},
\end{aligned}$$

we arrive at

$$P_1 \leq C \exp\Big( - \frac{M}{16} \lambda_{n,T} a_{n,T} + C \frac{\lambda_{n,T}^2}{nTh(nT)^\eta} \Big).$$

Moreover, choosing

$$r_{n,T} = \sqrt{\frac{nTh}{\tau_{n,T}^2 \log nT}},$$

we obtain that $\frac{\lambda_{n,T}^2}{nTh(nT)^\eta} = \log(nT)$ and $\lambda_{n,T} a_{n,T} = \log(nT)$. As a result,

$$P_1 \leq C \exp\Big(\Big[C - \frac{M}{16}\Big] \log nT\Big) \leq C(nT)^{-p},$$

where $p$ can be made arbitrarily large by choosing $M$ large enough. We next turn to $P_2$. Using (A.16), we obtain that

$$\begin{aligned} P_2 &\leq \sum_{j \neq i} \sum_{s=1}^{q_{n,T}} \mathbb{P}\Big(\big|\overline{B}_{j,2s-1}(x) - \overline{B}^*_{j,2s-1}(x)\big| > \frac{M a_{n,T}}{16 n q_{n,T}}\Big) \\ &\leq C \sum_{j \neq i} \sum_{s=1}^{q_{n,T}} \Big(\frac{\overline{C}_{n,T}}{a_{n,T}/nq_{n,T}}\Big)^{1/2} \alpha(r_{n,T}) \leq C(nT)^{-q}, \end{aligned}$$

where $q$ can be chosen arbitrarily large as the $\alpha$-coefficients decay exponentially fast.

Putting everything together, we arrive at

$$\begin{aligned} \mathbb{P}\Big(\max_{1 \leq l \leq L_{n,T}} \big|Q^{\leq}(x_l)\big| > \frac{M}{4} a_{n,T}\Big) &\leq \sum_{i=1}^{n} \sum_{l=1}^{L_{n,T}} \mathbb{P}\Big(\Big|\sum_{j \neq i} \sum_{t=1}^{T} I(\mathcal{C}_{iT}) \varphi_{it}(x_l) W_{jt}\Big| > \frac{M}{4} a_{n,T}\Big) \\ &\leq C n L_{n,T} \big[(nT)^{-p} + (nT)^{-q}\big]. \end{aligned}$$

If we choose the exponents $p$ and $q$ sufficiently large, then the right-hand side converges to zero at an arbitrarily fast polynomial rate. This completes the proof. $\square$

# Appendix B

# The Effect of Fragmentation in Trading on Market Quality in the UK Equity Market

## B.1 The regulatory framework under MiFID

The "Markets in Financial Instruments Directive (MiFID)" is a directive of the European Union that was adopted by the Council of the European Union and the European Parliament in April 2004 and became effective in November 2007. It replaces the "Investment Services Directive (ISD)" of 1993 that has become outdated by the fast speed of innovation in the financial industry. MiFID is the cornerstone of the "Financial Services Action Plan" that aims to foster the integration and harmonization of European financial markets. It provides a common regulatory framework for security markets across the 30 member states of the European Economic Area[1] to encourage the trading of securities and the provision of financial services across borders. The main pillars of MiFID are **market access**, **transparency** and **investor protection.**

1. **Market access.** MiFID abolished the monopoly position that many primary exchanges in the European Economic Area have had in equity trading. Under MiFID, orders can be executed on either regulated markets (RM), multilateral trading facilities (MTF) or systematic internalizers (SI). RMs and MTFs have similar trading functionalities but differ in the level of regulatory requirements. In contrast to MTFs, RMs must obtain authorization from a competent authority. While some MTFs have a visible (lit) order book, others operate as regulated dark pools. In a dark pool, traders submit their orders anonymously and they remain hidden until execution.[2] SIs are investment firms that execute client orders against other client orders or against their own inventories.

   The new entrants differentiate themselves on quality, price and technology that are usually tailored to speed-sensitive high frequency traders. In particular, MTF's typically adopt so-called maker-taker rebates that reward the provision of liquidity to the system, permit various types of orders and have small tick sizes. Additionally, their computer systems offer a lower latency when compared to regulated markets.

   While the number of RMs did not significantly increase after the introduction of MiFID, a large number of MTFs and SIs emerged in the post-MiFID period and successfully captured market share from the primary markets. At the end of October 2007, the European Securities and Markets Authority (ESMA) listed 93 RMs, 84 MTFs and 4 SIs. By the end of 2012, the number of MTFs had almost doubled to 151. While SIs are rare compared to MTFs, their number had grown to 13 by December 2012. In contrast, the number of RMs

[1] The European Economic Area consists of the 27 member states of the European Union as well as Norway, Iceland, and Liechtenstein.

[2] There are other, unregulated categories of dark pools that are registered as OTC venues or brokers (Gresse, 2012)

had only increased to 94.[3]

MiFID also extends the single passport concept that was already introduced in the ISD to establish a homogeneous European market governed by a common set of rules. The single passport concept enables investment firms that are authorized and regulated in their home state to serve customers in other EU member states.

2. **Transparency.** With an increasing level of fragmentation, information on prices and quantities available in the order books of different venues becomes dispersed. In response, MiFID introduced pre- and post-trade transparency provisions to enable investors to optimally decide where to execute their trade. Pre-trade transparency provisions apply to RMs and MTFs that operate a visible order book and require these venues to publish their order book in real time. Dark venues, OTC markets and SIs use waivers to circumvent the pre-trade transparency rules. To comply with post-trade transparency regulations, RMs, MTFs including regulated dark pools and OTC venues have to report executed trades to either the primary exchange or to a trade reporting facility (TRF) such as Markit BOAT.

3. **Investor protection.** MiFID introduces investor protection provisions to ensure that investment firms keep investors informed about their execution practises in a fragmented market place. An important part of these regulations is the best execution rule. Investment firms are required to execute orders that are on behalf of their clients at the best available conditions taking into account price, transaction costs, speed and likelihood of execution. Investment firms have to review their routing policy on a regular basis.

However, the financial crisis exposed several shortcomings of MiFID and the European Commission reacted to them by proposing a revision. The most important changes include the regulation of e.g. derivatives trading on “Organised Trading Facilities”, the introduction of safeguards for HFT, the improvement of transparency in equity, bonds and derivative markets, the reinforcement of supervisory powers in e.g. commodity markets and the strengthening of investor protection (European Commission (2011)).

## B.2 Trading venues

This appendix lists the individual trading venues that are used in our study.

[3] http://mifiddatabase.esma.europa.eu/, accessed on November 11, 2012

– **Lit venues:** Bats Europe, Chi-X, Equiduct, LSE, Nasdaq Europe, Nyse Arca, and Turquoise[4]

– **Regulated dark pools:** BlockCross, Instinet BlockMatch, Liquidnet, Nomura NX, Nyfix, Posit, Smartpool, and UBS MTF.

– **OTC venues:** Boat xoff, Chi-X OTC, Euronext OTC, LSE xoff, Plus, XOFF, and xplu/o.

– **Systematic internalizers:** Boat SI and London SI.

## B.3 System latency at the LSE

**Table C:** System latency at the LSE

| System | Implementation Date | Latency (Microseconds) |
|---|---|---|
| SETS | <2000 | 600000 |
| SETS1 | Nov 2001 | 250000 |
| SETS2 | Jan 2003 | 100000 |
| SETS3 | Oct 2005 | 55000 |
| TradElect | June 18, 2007 | 15000 |
| TradElect 2 | October 31, 2007 | 11000 |
| TradElect 3 | September 1, 2008 | 6000 |
| TradElect 4 | May 2, 2009 | 5000 |
| TradElect 4.1 | July 20, 2009 | 3700 |
| TradElect 5 | March 20, 2010 | 3000 |
| Millenium | February 14, 2011 | 113 |

Source: Brogaard et al. (2013) and own calculations.

## B.4 Econometric justification for quantile CCE estimation

We sketch an outline of the argument for the consistency of the quantile regression estimators used above. Harding and Lamarche (2013) consider the case with homogeneous panel data models; their theory does not apply to the heterogeneous model we study.

We consider a special case where we observe a sample of panel data $\{(Y_{it}, X_{it}) : i = 1, \ldots, n,\ t = 1, \ldots, T\}$. We first assume that the data come from the linear panel regression model

$$Y_{it} = \alpha_i + \beta_i X_{it} + \kappa_i f_t + \varepsilon_{it}$$

[4]On 21 December 2009, the London Stock Exchange Group agreed to take a 60% stake in trading platform Turquoise.

where $f_t$ denotes the unobserved common factor or factors. The covariates satisfy

$$X_{it} = \delta_i + \rho_i f_t + u_{it}$$

where in the Pesaran (2006) model the error terms satisfy the conditional moment restrictions $E(u_{it}^\intercal, \varepsilon_{it}|X_{it}, f_t) = 0$ with $u$ independent of $\varepsilon$. The unobserved factors $f_t$ are assumed to be either bounded and deterministic or a stationary ergodic sequence. Then assume that

$$\theta_i = \theta + \eta_i$$

where $\theta_i = (\alpha_i, \beta_i, \kappa_i, \delta_i, \rho_i)^\intercal$, $\theta = (\alpha, \beta, \kappa, \delta, \rho)^\intercal$ and $\eta_i$ are iid and independent of all the other random variables in the system. This is a special case of the model considered by Pesaran (2006). Letting $h_{0t} = \delta + \rho f_t$, we can write (provided $\rho \neq 0$)

$$Y_{it} = \alpha_i^* + \beta_i X_{it} + \kappa_i^* h_{0t} + \varepsilon_{it}$$

with $\alpha_i^* = \alpha_i - \delta\kappa_i/\rho$ and $\kappa_i^* = \kappa_i/\rho$, and note that $E(\varepsilon_{it}|X_{it}, h_{0t}) = 0$.

Taking cross-sectional averages we have

$$\overline{X}_t = \delta + \rho f_t + \overline{u}_t + \overline{\delta} - \delta + (\overline{\rho} - \rho) f_t = h_{0t} + O_p(n^{-1/2})$$

since $\overline{u}_t = O_p(n^{-1/2}) = \overline{\delta} - \delta = \overline{\rho} - \rho$. Therefore, we may consider the least squares estimator that minimizes $\sum_{t=1}^T \left\{Y_{it} - a - bX_{it} - c\overline{X}_t\right\}^2$ with respect to $\psi = (a, b, c)$, which yields a closed form estimator. This bears some similarities to the approach of Pesaran (2006) except that we do not include $\overline{Y}_t$ here Moon and Weidner (2015) advocate a QMLE approach, which would involve optimizing a pooled objective function over $\theta_i$, $i = 1, \ldots, n$ and $f_t$, $t = 1, \ldots, T$. In the QMLE case this may be feasible, but in the case with more nonlinearity such as quantiles as below this seems infeasible.

We now turn to quantile regression, and in particular median regression. We shall now assume that $\mathrm{med}(\varepsilon_{it}|X_{it}, f_t) = 0$ and maintain the assumptions that $E(u_{it}) = 0$ with $u$ independent of $\varepsilon$, so that $\overline{X}_t = \delta + \rho f_t + \overline{u}_t = h_{0t} + O_p(n^{-1/2})$ as before. We consider a more general class of estimators based on minimizing the objective function

$$Q_{Ti}(\psi) = \frac{1}{T}\sum_{t=1}^T \lambda(Y_{it} - a - bX_{it} - c\overline{X}_t)$$

over $\psi$, where $\lambda(t) = |t|$. The approximate first order conditions are based on

$$M_{Ti}(\psi; \overline{X}_1, \ldots, \overline{X}_T) = \frac{1}{T}\sum_{t=1}^T \begin{pmatrix} 1 \\ X_{it} \\ \overline{X}_t \end{pmatrix} \mathrm{sign}\left(Y_{it} - \alpha - \beta X_{it} - \gamma\overline{X}_t\right)$$

$$= \frac{1}{T}\sum_{t=1}^{T} m_{it}(\psi, \overline{X}_t)$$

We discuss now the properties of $\widehat{\psi}_i$, the zero of $M_{Ti}(\psi; \overline{X}_1, \ldots, \overline{X}_T)$. For this purpose we can view $\widehat{\psi}_i$ as an example of a semiparametric estimator as considered in Chen et al. (2003). That is, $\overline{X}_t$ is a preliminary estimator of the "function" $h_{0t} = \delta + \rho f_t$.

An important part of the argument is to show the uniform consistency of this estimate

$$\max_{1\le t\le T} \left|\overline{X}_t - \delta - bf_t\right| \le \max_{1\le t\le T} |\overline{u}_t| + \left|\overline{\delta} - \delta\right| + (\max_{1\le t\le T} |f_t|)\, |\overline{\rho} - \rho| = o_p(1).$$

By elementary arguments we have $\max_{1\le t\le T} |\overline{u}_t| = o_p(T^{\kappa} n^{-1/2})$ for some $\kappa$ depending on the number of moments that $u_{it}$ possesses. Similarly, $\max_{1\le t\le T} |f_t| = O_p(T^{\kappa})$ under the same moment conditions.

For compactness, let us denote $M_{Ti}(\psi; \overline{X}_1, \ldots, \overline{X}_T)$ by $M_{Ti}(\psi, \widehat{h})$, where $\widehat{h} = (\overline{X}_1, \ldots, \overline{X}_T)$. The approach of CLV is to approximate the estimator

$$\widehat{\psi} = \arg\min_{\psi\in\Psi} ||M_{Ti}(\psi, \widehat{h})||$$

by the estimator

$$\overline{\psi} = \arg\min_{\psi\in\Psi} ||M_{Ti}(\theta, h_0)||$$

where $h_0 = (h_{01}, \ldots, h_{0T})$ is the true sequence. In the case where $m_{it}(\psi, h)$ is smooth in $h$, this follows by straightforward Taylor expansion and using the uniform convergence result above. In the quantile case, some empirical process techniques are needed as usual, but they are standard. The estimator $\overline{\psi}$ is just the standard quantile regression estimator of the parameters in the case where $h_{0t}$ is observed and so consistency follows more or less by a standard route, namely, the strong law of large numbers implies that

$$\begin{aligned} M_{Ti}(\psi, h_0) &= \frac{1}{T}\sum_{t=1}^{T} \begin{pmatrix} 1 \\ X_{it} \\ \delta + \rho f_t \end{pmatrix} \operatorname{sign}\left(Y_{it} - \alpha - \beta X_{it} - \gamma\delta - \rho\gamma f_t\right) \\ &\to E_i\left[\begin{pmatrix} 1 \\ X_{it} \\ \delta + \rho f_t \end{pmatrix} \operatorname{sign}\left(Y_{it} - \alpha - \beta X_{it} - \gamma(\delta + \rho f_t)\right)\right] \\ &\equiv M_i(\psi) \end{aligned}$$

which is uniquely minimized at the true value of $\psi$. Here, $E_i$ means expectation conditional on $\psi_i$.

In fact, because of the independence of $u, \varepsilon$, the joint distribution of $\varepsilon_{it}, X_{it}, f_t$

factors into the product of the conditional distribution of $\varepsilon_{it}|f_t$ the conditional distribution of $u_{it}|f_t$ and the marginal distribution of $f_t$. We calculate $M_i(\psi)$. We have

$$\begin{aligned} M_{1i}(\psi) =& E_i\left[\text{sign}\left(Y_{it} - \alpha - \beta X_{it} - \gamma\delta - \rho\gamma f_t\right)\right] \\ =& \int [1 - 2G((\alpha_i - \alpha) + (\beta_i - \beta)(u + \delta_i + \rho_i f) \\ &+ (\gamma_i - \gamma)(\delta + \rho f)|f)] r(u|f) q(f) d\varepsilon du df \end{aligned}$$

where $G$ is the c.d.f of $\varepsilon|f$ with density $g$ and $r$ is the density of $u|f$ and $q$ is the marginal density of $f$. It follows that $M_{1i}(\psi_0) = 0$ by the conditional median restriction. Similarly with $M_{ji}(\psi)$, $j = 2, 3$. Under some conditions can establish the uniqueness needed for consistency. We can further calculate $\partial M_{1i}(\psi)/\partial\psi$.

The next question is whether the estimation of $h_0$ by $\widehat{h}$ affects the limiting distribution. In this case we consider the sequence $h^* = (h_1^*, \ldots, h_T^*)$

$$\begin{aligned} E_i\left[m_{it}(\psi, h_t^*)|f_t\right] =& E_i\left[m_{it}(\psi, h_{0t})|f_t\right] + \frac{\partial}{\partial h} E_i\left[m_{it}(\psi, h_{0t})|f_t\right]\left[h_t^* - h_{0t}\right] \\ &+ \frac{\partial^2}{\partial h^2} E_i\left[m_{it}(\psi, \overline{h}_t)|f_t\right]\left[h_t^* - h_{0t}\right]^2 \end{aligned}$$

for intermediate values $\overline{h}_t$. Then we can show that $\partial E_i\left[m_{it}(\psi, h_{0t})|f_t\right]/\partial h$ has a finite expectation and so

$$\begin{aligned} & \frac{1}{T}\sum_{t=1}^{T} \frac{\partial}{\partial h} E_i\left[m_{it}(\psi_0, h_{0t})|f_t\right]\left[\widehat{h}_t - h_{0t}\right] \\ = & \frac{1}{T}\sum_{t=1}^{T} \frac{\partial}{\partial h} E_i\left[m_{it}(\psi_0, h_{0t})|f_t\right]\left[\overline{u}_t + \overline{\delta} - \delta + (\overline{\rho} - \rho) f_t\right] = O_p(n^{-1/2}T^{-1/2}) \end{aligned}$$

because $E_i\left[\overline{u}_t + \overline{\delta} - \delta + (\overline{\rho} - \rho) f_t | f_t\right] = 0$. Furthermore,

$$\begin{aligned} & \frac{1}{T}\sum_{t=1}^{T} \frac{\partial^2}{\partial h^2} E\left[m_{it}(\psi, \overline{h}_t)|f_t\right]\left[\widehat{h}_t - h_{0t}\right]^2 \\ = & \frac{1}{T}\sum_{t=1}^{T} \frac{\partial^2}{\partial h^2} E\left[m_{it}(\psi, \overline{h}_t)|f_t\right]\left[\overline{u}_t + \overline{\delta} - \delta + (\overline{\rho} - \rho) f_t\right]^2 = O_p(n^{-1}) \end{aligned}$$

so that we need $T/n^2 \to 0$. It follows that the limiting distribution is the same as that of $\overline{\psi}$. The conditions of CLV Theorem 1 and 2 are satisfied. In particular, for:

$$\Gamma_1(\psi, h_o) = \frac{\partial}{\partial \psi} M(\psi) = -2 \times p \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \begin{pmatrix} 1 & X_{it} & h_{0t} \\ X_{it} & X_{it}^2 & X_{it}h_{0t} \\ h_{0t} & X_{it}h_{0t} & h_{0t}^2 \end{pmatrix} g(0|X_{it}, f_t)$$

$$
\begin{aligned}
V_1 &= \mathrm{var}[m_{it}(\psi_0, h_{0t}))] \\
&= \left(\begin{array}{ccc}
1 & \delta_i + \rho_i E f_t & \delta + \rho E f_t \\
\delta_i + \rho_i E f_t & \sigma_u^2 + \delta_i^2 + \rho_i^2 E f_t^2 + 2\delta_i \rho_i E f_t & \delta_i \delta + \delta_i \rho E f_t^2 + (\delta_i \rho + \delta \rho_i) E f_t \\
\delta + \rho E f_t & \delta_i \delta + \rho_i \rho E f_t^2 + (\delta_i \rho + \delta \rho_i) E f_t & \delta^2 + \rho^2 E f_t^2 + 2\delta \rho E f_t
\end{array}\right)
\end{aligned}
$$

we have

$$\sqrt{T}(\widehat{\psi}_i - \psi_i) \Longrightarrow \mathcal{N}[0, \Omega], \text{ where } \Omega = (\Gamma_1^\intercal \Gamma_1)^{-1} \Gamma_1^\intercal V_1 \Gamma_1 (\Gamma_1^\intercal \Gamma_1)^{-1}.$$

It follows that for each $i$

$$\sqrt{T}(\widehat{\beta}_i - \beta_i) \xrightarrow{d} N(0, \Omega_{\beta\beta i})$$

where $\Omega_{\beta\beta i}$ is the appropriate submatrix of above.

In the case that $g(0|X_{it}, f_t) = g(0)$ we have

$$\Omega_i = \frac{1}{4g(0)} V_1^{-1}.$$

Under some additional conditions we may obtain the asymptotic behaviour of the mean group estimator $\widehat{\beta} = n^{-1} \sum_{i=1}^n \widehat{\beta}_i$. Specifically, we have

$$\sqrt{n}(\widehat{\beta} - \beta) \xrightarrow{d} N(0, \Sigma_{\beta\beta})$$

where $\Sigma_{\beta\beta} = \mathrm{var}(v_{\beta i})$. This follows because

$$
\begin{aligned}
\widehat{\beta} - \beta &= \frac{1}{n} \sum_{i=1}^n (\widehat{\beta}_i - \beta_i) + \frac{1}{n} \sum_{i=1}^n (\beta_i - \beta) \\
&= \frac{1}{n} \sum_{i=1}^n v_{\beta i} + O_p(T^{-1/2} n^{-1/2}) + O_p(n^{-1}),
\end{aligned}
$$

because the averaging over $i$ reduces the orders, for example

$$\frac{1}{n} \sum_{i=1}^n \frac{1}{T} \sum_{t=1}^T \begin{pmatrix} 1 \\ X_{it} \\ h_{0t} \end{pmatrix} \mathrm{sign}\,(\varepsilon_{it}) = O_p(T^{-1/2} n^{-1/2}).$$

The argument extends to the more general specification considered in this chapter.

## B.5 Robustness

### Alternative measures of market quality

Measuring market quality is inherently difficult, and there is an ongoing debate on what constitutes a good measure of market quality. In view of this controversy, this section investigates the robustness of the results in chapter 2 to a variety of alternative measures of market quality. The particular measures we consider are total (Parkinson) volatility, idiosyncratic volatility, within day and overnight volatility, efficiency, and Amihud illiquidity.

*Volatility.* In the main paper, total volatility is measured by the Rogers-Satchell estimator. An alternative measure is due to Parkinson (1980).[5] The Parkinson estimator of total volatility can be computed as

$$V_{it_j}^{P} = \frac{1}{4\ln 2}\left(\ln P_{it_j}^{H} - \ln P_{it_j}^{L}\right)^2 .$$

Figure B.1 documents that the Parkinson volatility estimator is highly correlated with the Rogers-Satchell estimator.

Some have argued that HFT activity and the associated market fragmentation leads to higher volatility through the endogenous trading risk process (Foresight (2012)). Therefore, we also obtained measures of overnight volatility that reflect changes in prices that occur between the closing auction and the opening auction and are therefore not subject to the influence of the continuous trading process. In particular, we decompose volatility into overnight volatility and intraday volatility,

$$\begin{aligned} V_{it_j}^{day} &= (\ln P_{it_j}^{C} - \ln P_{it_j}^{O})^2 \\ V_{it_j}^{night} &= (\ln P_{it_j}^{O} - \ln P_{it-1_j}^{C})^2 . \end{aligned}$$

Unfortunately, we can't completely separate out the auction component and the continuous trading component, which would also be of interest. Figure B.2 reports the time series of the cross-sectional quantiles of (the log of) overnight and within day volatility, as well as their ratio. The two series move quite closely together. There is an increase during the early part of the series followed by a decrease later, as with total volatility. The ratio of the two series shows no discernible trend at any quantile over this period. It seems that volatility increases and decreases but in no sense has become concentrated intraday relative to overnight.

In addition, we computed a measure of idiosyncratic volatility. In principle, idiosyncratic risk is diversifiable and should not be rewarded in terms of expected returns. We consider whether the effects of fragmentation take place on volatility

[5] We also measured total volatility by the simple range estimator $V_{it_j} = \frac{P_{it_j}^{H} - P_{it_j}^{L}}{P_{it_j}^{L}}$. The results for this estimator are very similar to the Parkinson estimator and are available upon request.

through the common or idiosyncratic part. If it is on the idiosyncratic component of returns then it should have less impact on diversified investors, i.e., big funds and institutions. Specifically, idiosyncratic volatility is calculated as the squared residuals from a regression of individual close-to-close returns on index close-to-close returns. Common volatility is then obtained as the square of the slope coefficient multiplied by the variance of the index return. Cross-sectional quantiles of idiosyncratic and common volatility are shown in Figure B.3. The sharp increase in volatility during the financial crisis is more pronounced for the common component.

*Liquidity.* This appendix considers an alternative measures of liquidity based on daily transaction data. In particular, we use the Amihud (2002) measure that is defined as

$$IL_{it_j} = \frac{|R_{it_j}|}{Vol_{it_j}},$$

where $Vol_{it_j}$ is the daily turnover and $R_{it_j}$ are daily close to close returns. Goyenko et al. (2009) argue that the Amihud measure provides a good proxy for the price impact. Figure B.4 compared the cross-sectional quantiles of the Amihud measure and bid-ask spreads. The two measures seem to move quite closely together and share a similar trajectory with volatility measures. Towards the end of the sample there does seem to be a narrowing of the cross sectional distribution of bid ask spreads.

*Efficiency.* A market that is grossly "inefficient" would be indicative of poor market quality. Hendershott (2011) gives a discussion of market efficiency and how it can be interpreted in a high frequency world. We shall take a rather simple approach and base our measure of inefficiency/predictability on just the daily closing price series (weak form) and confine our attention to linear methods. In this world, efficiency or lack thereof, can be measured by the degree of autocorrelation in the stock return series. We compute an estimate of the weekly lag one autocorrelation denoted by $\rho_{it}(k) = \mathrm{corr}(R_{it_j}, R_{it_{j-k}})$, $k = 1, 2$, where $R_{t_j}$ denotes the close to close return for stock $i$ on day $j$ within week $t$; the variance and covariance are computed with daily data within week $t$. Under the efficient markets hypothesis this quantity should be zero, but in practice this quantity is different from zero and sometimes statistically significantly different from zero. Since the series is computed from at most five observations it is quite noisy, we use the small sample adjustment from Campbell et al. (2012), eq. 2.4.13)

$$\widehat{\rho}_{it}^{A} = \widehat{\rho}_{it} + \frac{1}{N_{it} - 1}[1 - \widehat{\rho}_{it}^{2}],$$

where $\widehat{\rho}_{it}$ is the sample autocorrelation based on $N_{it} \leq 5$ daily observations. In this case, $\widehat{\rho}_{it}^{A}$ is an approximately unbiased estimator of weekly efficiency. We take the absolute value of the efficiency measure. Figure B.5 reports cross-sectional quantiles of our efficiency measure. The median inefficiency is around 0.3.[6] The variation of the efficiency measures over time does not suggest that the efficiency of daily stock returns either improves or worsens over this time period.

Our finding that visible fragmentation and dark trading have a negative effect on total and temporary volatility is robust to using alternative measures of volatility such as Parkinson or within-day volatility (Tables B.1 - B.2). If we measure market quality by the Amihud (2002) illiquidity measure, we find that a higher degree of overall or visible fragmentation is associated with less liquid markets. Dark trading is found to improve liquidity. Because the Amihud (2002) liquidity measure is closely related to LSE volume, these results probably in part reflect our findings for LSE volume in the main paper. For efficiency, we cannot find significant effects.

Turning to the effect of fragmentation on the variability of market quality (Tables B.3-B.4), we find that dark trading increases the variability of total (Parkinson) volatility which is consistent with the results reported in chapter 2. We also document that a higher level of overall fragmentation reduces the variability of Amihud illiquidity.

## FTSE 100 and FTSE 250 subsamples

In chapter 2, we only report results for a pooled sample of all FTSE 100 and 250 firms. In this appendix, we complement our main results by splitting the sample into FTSE 100 and FTSE 250 stocks. The FTSE 100 index is composed of the 100 largest firms listed on the LSE according to market capitalization, while the FTSE 250 index comprises the "mid-cap" stocks.

When comparing the effect of market fragmentation on market quality for FTSE 100 and FTSE 250 firms, interesting differences emerge: the effects of overall fragmentation on temporary volatility and global volume can be attributed to FTSE 100 firms (Tables B.5-B.6). The negative effect of dark trading on volatility is only observed for FTSE 250 firms (Tables B.7-B.8). That effect is even positive for FTSE 100 firms. But in contrast with FTSE 250 firms, visible fragmentation is associated with lower volatility for FTSE 100 firms. Inspecting the results for the variability of market quality, overall fragmentation reduces the variability of LSE trading volume only for FTSE 250 firms, while dark trading increases the variability of LSE volumes for FTSE 100 firms (Tables B.9 - B.12).

[6]Note that when $\widehat{\rho}_{it} = 0$, $\widehat{\rho}_{it}^{A} = 0.25$ because $N_{it} = 5$ most of the time. Therefore, the bias adjusted level is quite high.

## Omitting common factors

Related research uses panel data specifications as for example fixed effects estimators that cannot account for unobserved common factors (Gresse (2011), De Jong et al. (2015)). To address concerns about endogeneity of fragmentation and dark trading, some papers use instrumental variable methods. But they do not however instrument other included covariates which are likely to be jointly determined along with the outcome variable. Specifically, some include volume and volatility as exogenous covariates in equations for market quality measures. In contrast, the CCE methodology used in this chapter can control for common unobserved factors that affect both dependent and independent variables.

To illustrate the importance of controlling for unobserved common factors, we re-estimate our results using a heterogeneous panel data model without common factors. This model can be obtained as a special case of our econometric model where $d_t$ is a vector of ones and there are no unobserved common factors $f_t$. A version of this model with homogenous coefficients has been used by Gresse (2011), among others. As reported in Table B.13, omitting observed and unobserved common factors leads to results that differ in magnitude and statistical significance with the exception of LSE volume. However, the large increase in our measure of cross-sectional dependence (CSD) indicates that this model is misspecified because unobserved common shocks such as changes in trading technology or high frequency trading are omitted that are likely to affect both market quality and fragmentation.

## Stochastic Dominance

Finally, we investigate if the distribution of market quality under competition stochastically dominates its distribution in a monopolistic market using the method in Linton et al. (2005). If market quality is measured by bid-ask spreads, we find evidence of second order stochastic dominance of competition over monopoly, and vice versa for volatility. However, this evidence is only indicative as we did not formally obtain critical values for the test statistic.

**Table B.1:** The effect of fragmentation on market quality for alternative measures of market quality

| | Total (Parkinson) volatility | Idiosync. volatility | Daily volatility | Overnight volatility | Efficiency | Illiquidity |
|---|---|---|---|---|---|---|
| Constant | -7.713 | -6.987 | -5.507 | -14.926 | 0.562 | -13.652 |
| | (-8.817) | (-4.855) | (-3.025) | (-10.13) | (2.738) | (-14.019) |
| Frag. | 0.208 | 0.416 | -0.11 | -1.916 | -0.025 | -0.524 |
| | (0.383) | (0.518) | (-0.134) | (-1.919) | (-0.23) | (-1.112) |
| Frag. sq. | -0.534 | -0.988 | -0.368 | 1.1 | 0.056 | 1.341 |
| | (-1.269) | (-1.446) | (-0.55) | (1.356) | (0.579) | (3.315) |
| Market cap. | -0.499 | -0.48 | -0.591 | -0.48 | -0.039 | -0.322 |
| | (-6.936) | (-3.694) | (-5.561) | (-4.238) | (-2.539) | (-4.528) |
| ME (frag.) | -0.349 | -0.615 | -0.495 | -0.768 | 0.033 | 0.875 |
| | (-2.634) | (-3.146) | (-2.478) | (-3.43) | (1.303) | (8.422) |
| $\Delta_{Frag.}$ | -0.238 | -0.408 | -0.418 | -0.998 | 0.021 | 0.595 |
| | (-1.154) | (-1.457) | (-1.402) | (-2.821) | (0.592) | (3.797) |
| Adjusted $R^2$ | 0.755 | 0.41 | 0.419 | 0.442 | 0.022 | 0.866 |

*Notes*: Coefficients are quantile CCE mean group estimates. t-statistics are shown in parenthesis. Market capitalisation and dependent variables (except of idiosyncratic volatility and efficiency) are in logs. Lagged index return, VIX and Christmas and New Year effects are included as observable factors. $\Delta_{Frag.}$ is defined as $\widehat{\beta}_1 + \widehat{\beta}_2(H + L)$ and evaluated at $H = \max(\text{Frag.})$ and $L = \min(\text{Frag.})$. ME denotes marginal effects. The adjusted $R^2$ is the $R^2$ calculated from pooling the individual total and residual sums of squares, adjusted for the number of regressors.

**Table B.2:** The effects of visible fragmentation and dark trading on market quality for alternative measures of market quality

| | Total (Parkinson) volatility | Idiosync. volatility | Daily volatility | Overnight volatility | Efficiency | Illiquidity |
|---|---|---|---|---|---|---|
| Constant | -7.061 | -7.039 | -3.303 | -14.786 | 0.348 | -12.065 |
| | (-8.882) | (-4.277) | (-2.046) | (-9.409) | (1.423) | (-12.319) |
| Vis. frag. | 0.263 | -1.023 | -0.797 | 0.04 | 0.019 | -0.249 |
| | (0.934) | (-1.878) | (-1.697) | (0.081) | (0.238) | (-0.506) |
| Vis. frag. sq. | -0.815 | 0.361 | 0.04 | -0.422 | -0.011 | 0.873 |
| | (-2.472) | (0.547) | (0.066) | (-0.672) | (-0.106) | (1.631) |
| Dark | 0.061 | -0.237 | 0.98 | -1.033 | 0.046 | -0.752 |
| | (0.264) | (-0.482) | (1.877) | (-2.467) | (0.59) | (-3.023) |
| Dark sq. | -0.202 | 0.367 | -1.398 | 1.125 | -0.031 | -0.096 |
| | (-0.858) | (0.757) | (-2.749) | (2.555) | (-0.384) | (-0.397) |
| Market cap. | -0.405 | -0.441 | -0.497 | -0.3 | -0.04 | -0.217 |
| | (-5.698) | (-3.066) | (-4.329) | (-2.447) | (-2.228) | (-2.989) |
| ME (Vis. frag) | -0.313 | -0.768 | -0.769 | -0.258 | 0.011 | 0.368 |
| | (-2.99) | (-4.004) | (-4.029) | (-1.23) | (0.394) | (2.058) |
| ME (Dark) | -0.124 | 0.1 | -0.303 | 0 | 0.018 | -0.84 |
| | (-1.891) | (0.585) | (-1.991) | (0.004) | (0.746) | (-9.526) |
| $\Delta_{Vis.frag.}$ | -0.306 | -0.771 | -0.769 | -0.255 | 0.011 | 0.361 |
| | (-2.899) | (-3.991) | (-4.029) | (-1.211) | (0.396) | (1.99) |
| $\Delta_{Dark}$ | -0.14 | 0.129 | -0.417 | 0.092 | 0.015 | -0.848 |
| | (-2.111) | (0.758) | (-2.804) | (0.721) | (0.62) | (-9.679) |
| Adjusted $R^2$ | 0.773 | 0.417 | 0.429 | 0.455 | 0.031 | 0.871 |

*Notes*: See Table B.1 except that $X = \{\text{Vis. frag, Dark}\}$.

**Table B.3:** The effect of fragmentation on the variability of market quality for alternative measures of market quality (conditional variance model)

| | Total (Parkinson) volatility | Idiosync. volatility | Daily volatility | Overnight volatility | Efficiency | Illiquidity |
|---|---|---|---|---|---|---|
| Constant | -0.091 | 1.119 | -0.38 | -1.47 | 0.097 | 0.949 |
| | (-0.366) | (0.871) | (-0.421) | (-1.503) | (3.753) | (2.679) |
| Frag. | 0.015 | -0.234 | -0.671 | -0.004 | -0.031 | -0.48 |
| | (0.154) | (-0.418) | (-1.413) | (-0.007) | (-2.057) | (-2.377) |
| Frag. sq. | -0.015 | 0.178 | 0.708 | 0.04 | 0.031 | 0.404 |
| | (-0.158) | (0.343) | (1.681) | (0.08) | (2.391) | (2.251) |
| Market cap. | -0.008 | -0.152 | -0.001 | 0.088 | -0.003 | -0.023 |
| | (-0.366) | (-1.663) | (-0.018) | (1.052) | (-1.506) | (-0.915) |
| ME (frag.) | 0 | -0.048 | 0.068 | 0.038 | 0.001 | -0.058 |
| | (-0.003) | (-0.4) | (0.59) | (0.225) | (0.172) | (-1.175) |
| $\Delta_{Frag.}$ | 0.003 | -0.085 | -0.08 | 0.029 | -0.006 | -0.143 |
| | (0.095) | (-0.516) | (-0.509) | (0.139) | (-1.034) | (-2.125) |
| Adjusted $R^2$ | 0.002 | -0.04 | -0.084 | -0.068 | -0.088 | -0.004 |

*Notes*: Dependent variables are squared median regression residuals. Coefficients are quantile CCE mean group estimates. t-statistics are shown in parenthesis. Market capitalization is in logs. Lagged index return, VIX and Christmas and New Year effects are included as observable factors. $\Delta_{Frag.}$ is defined as $\widehat{\beta} + \widehat{\gamma}(H + L)$ and evaluated at $H = \max(\text{Frag.})$ and $L = \min(\text{Frag.})$. ME denotes marginal effects. The adjusted $R^2$ is the $R^2$ calculated from pooling the individual total and residual sums of squares, adjusted for the number of regressors.

**Table B.4:** The effect of visible fragmentation and dark trading on the variability of market quality for alternative measures of market quality (conditional variance model)

| | Total (Parkinson) volatility | Idiosync. volatility | Daily volatility | Overnight volatility | Efficiency | Illiquidity |
|---|---|---|---|---|---|---|
| Constant | -0.356 | 2.445 | 0.863 | -2.094 | 0.089 | 0.547 |
| | (-1.383) | (1.88) | (0.834) | (-2.168) | (2.686) | (1.54) |
| Vis. frag. | -0.165 | -1.724 | -2.016 | 0.268 | 0.005 | -0.379 |
| | (-1.374) | (-1.321) | (-2.447) | (0.747) | (0.482) | (-3.733) |
| Vis. frag. sq. | 0.17 | 1.433 | 2.382 | -0.299 | 0.001 | 0.591 |
| | (1.219) | (1.213) | (2.985) | (-0.598) | (0.054) | (3.535) |
| Dark | 0.025 | -0.396 | -0.65 | -0.838 | -0.017 | -0.243 |
| | (0.362) | (-0.963) | (-1.683) | (-2.827) | (-2.129) | (-2.465) |
| Dark sq. | 0.056 | 0.544 | 0.711 | 0.927 | 0.022 | 0.257 |
| | (0.775) | (1.356) | (1.825) | (2.757) | (2.453) | (2.671) |
| Market cap. | -0.005 | -0.104 | -0.026 | -0.083 | 0 | 0.007 |
| | (-0.253) | (-1.086) | (-0.328) | (-0.949) | (-0.074) | (0.274) |
| ME (Vis. frag) | -0.045 | -0.711 | -0.333 | 0.057 | 0.005 | 0.039 |
| | (-1.009) | (-1.394) | (-0.984) | (0.376) | (1.361) | (0.605) |
| ME (Dark) | 0.076 | 0.104 | 0.003 | 0.013 | 0.003 | -0.008 |
| | (3.447) | (0.784) | (0.025) | (0.149) | (1.046) | (-0.256) |
| $\Delta_{Vis.frag.}$ | -0.047 | -0.724 | -0.354 | 0.059 | 0.005 | 0.033 |
| | (-1.033) | (-1.394) | (-1.031) | (0.395) | (1.355) | (0.531) |
| $\Delta_{Dark}$ | 0.081 | 0.148 | 0.061 | 0.088 | 0.005 | 0.013 |
| | (3.457) | (1.129) | (0.507) | (0.898) | (1.588) | (0.452) |
| Adjusted $R^2$ | -0.026 | -0.027 | -0.074 | -0.064 | -0.075 | -0.037 |

*Notes*: See Table B.3 except that $X = \{\text{Vis. frag, Dark}\}$.

**Table B.5:** The effect of fragmentation on market quality for FTSE 100 firms

| | Total volatility | Temp. volatility | BA spreads | Global volume | LSE volume |
|---|---|---|---|---|---|
| Constant | -2.74 | -8.643 | 9.955 | 1.286 | 3.546 |
| | (-2.296) | (-10.29) | (5.771) | (1.032) | (3.332) |
| Frag. | 1.141 | -2.935 | -0.02 | 1.711 | 2.197 |
| | (1.181) | (-3.147) | (-0.035) | (3.076) | (4.326) |
| Frag. sq. | -1.216 | 2.365 | 0.184 | -1.232 | -3.115 |
| | (-1.616) | (3.252) | (0.38) | (-2.457) | (-7.203) |
| Market cap. | -0.44 | -0.38 | -0.335 | -0.533 | -0.52 |
| | (-3.857) | (-4.993) | (-2.952) | (-6.469) | (-6.71) |
| ME (frag.) | -0.501 | 0.26 | 0.229 | 0.046 | -2.012 |
| | (-2.403) | (1.36) | (1.417) | (0.269) | (-15.503) |
| $\Delta_{Frag.}$ | 0.087 | -0.883 | 0.14 | 0.642 | -0.506 |
| | (0.245) | (-2.627) | (0.752) | (4.153) | (-3.223) |
| Adjusted $R^2$ | 0.777 | 0.173 | 0.605 | 0.801 | 0.831 |

*Notes*: See Table 2.1.

**Table B.6:** The effect of fragmentation on market quality for FTSE 250 firms

| | Total volatility | Temp. volatility | BA spreads | Global volume | LSE volume |
|---|---|---|---|---|---|
| Constant | -8.503 | -10.327 | 3.584 | 2.195 | 2.18 |
| | (-8.268) | (-13.225) | (3.743) | (2.639) | (2.336) |
| Frag. | -0.193 | -0.16 | 0.072 | -0.658 | -0.276 |
| | (-0.282) | (-0.316) | (0.258) | (-1.876) | (-0.837) |
| Frag. sq. | -0.162 | 0.012 | -0.164 | 0.707 | -1.091 |
| | (-0.297) | (0.029) | (-0.651) | (2.012) | (-3.298) |
| Market cap. | -0.437 | -0.293 | -0.326 | -0.058 | -0.084 |
| | (-4.379) | (-4.599) | (-3.772) | (-0.682) | (-0.979) |
| ME (frag.) | -0.359 | -0.148 | -0.096 | 0.064 | -1.392 |
| | (-2.102) | (-1.296) | (-1.258) | (0.635) | (-14.205) |
| $\Delta_{Frag.}$ | -0.328 | -0.15 | -0.065 | -0.069 | -1.186 |
| | (-1.301) | (-0.852) | (-0.682) | (-0.635) | (-11.469) |
| Adjusted $R^2$ | 0.713 | 0.094 | 0.706 | 0.738 | 0.714 |

*Notes*: See Table 2.1.

**Table B.7:** The effects of visible fragmentation and dark trading on market quality for FTSE 100 firms

| | Total volatility | Temp. volatility | BA spreads | Global volume | LSE volume |
|---|---|---|---|---|---|
| Constant | -2.643 | -7.637 | 8.131 | 4.08 | 5.067 |
| | (-1.852) | (-7.171) | (4.587) | (4.744) | (5.14) |
| Vis. frag. | -0.3 | -4.244 | 0.221 | -0.87 | -0.734 |
| | (-0.445) | (-8.073) | (0.628) | (-2.12) | (-1.825) |
| Vis. frag. sq. | -0.597 | 4.121 | 0.001 | 0.916 | -0.679 |
| | (-0.903) | (7.412) | (0.002) | (2.015) | (-1.498) |
| Dark | -0.003 | 1.217 | 0.052 | 0.98 | 0.864 |
| | (-0.009) | (3.507) | (0.14) | (3.189) | (2.185) |
| Dark sq. | 0.315 | -1.395 | -0.015 | 1.504 | 0.546 |
| | (0.676) | (-3.213) | (-0.037) | (4.333) | (1.269) |
| Market cap. | -0.332 | -0.29 | -0.326 | -0.46 | -0.47 |
| | (-2.539) | (-3.069) | (-3.094) | (-5.995) | (-5.859) |
| ME (vis. frag) | -0.91 | -0.028 | 0.222 | 0.068 | -1.428 |
| | (-4.674) | (-0.16) | (1.12) | (0.525) | (-10.759) |
| ME (dark) | 0.234 | 0.165 | 0.041 | 2.114 | 1.275 |
| | (2.098) | (1.668) | (0.329) | (21.692) | (9.818) |
| $\Delta_{Vis.frag.}$ | -0.715 | -1.378 | 0.222 | -0.233 | -1.206 |
| | (-2.67) | (-6.945) | (1.585) | (-1.731) | (-9.234) |
| $\Delta_{Dark}$ | 0.303 | -0.139 | 0.038 | 2.442 | 1.394 |
| | (2.02) | (-1.041) | (0.321) | (23.905) | (11.101) |
| Adjusted $R^2$ | 0.784 | 0.193 | 0.617 | 0.846 | 0.848 |

*Notes*: See Table 2.2.

**Table B.8:** The effects of visible fragmentation and dark trading on market quality for FTSE 250 firms

| | Total volatility | Temp. volatility | BA spreads | Global volume | LSE volume |
|---|---|---|---|---|---|
| Constant | -9.696 | -11.53 | 0.588 | 1.368 | 3.05 |
| | (-9.159) | (-12.407) | (0.465) | (1.692) | (3.456) |
| Vis. frag. | 1.277 | 0.839 | 0.565 | 0.334 | 0.03 |
| | (3.855) | (3.419) | (2.107) | (1.511) | (0.115) |
| Vis. frag. sq. | -1.969 | -1.164 | -0.787 | -1.035 | -1.706 |
| | (-4.665) | (-3.574) | (-2.222) | (-3.561) | (-5.192) |
| Dark | -0.531 | 0.032 | -0.42 | -0.071 | -0.073 |
| | (-1.775) | (0.121) | (-1.446) | (-0.275) | (-0.28) |
| Dark sq. | 0.221 | -0.325 | 0.297 | 1.972 | 1.312 |
| | (0.879) | (-1.403) | (1.137) | (9.367) | (5.59) |
| Market cap. | -0.487 | -0.371 | -0.318 | -0.343 | -0.311 |
| | (-5.184) | (-5.328) | (-3.531) | (-4.021) | (-3.494) |
| ME (vis. frag) | 0.031 | 0.102 | 0.067 | -0.321 | -1.05 |
| | (0.223) | (0.98) | (0.654) | (-2.879) | (-9.37) |
| ME (dark) | -0.308 | -0.295 | -0.121 | 1.916 | 1.25 |
| | (-4.202) | (-4.625) | (-1.644) | (26.542) | (18.581) |
| $\Delta_{Vis.frag.}$ | -0.097 | 0.026 | 0.015 | -0.389 | -1.161 |
| | (-0.728) | (0.253) | (0.155) | (-3.472) | (-10.722) |
| $\Delta_{Dark}$ | -0.31 | -0.292 | -0.123 | 1.899 | 1.238 |
| | (-4.162) | (-4.519) | (-1.665) | (25.949) | (18.291) |
| Adjusted $R^2$ | 0.735 | 0.114 | 0.671 | 0.831 | 0.764 |

*Notes*: See Table 2.1.

**Table B.9:** The effect of fragmentation on the variability of market quality for FTSE 100 firms (conditional variance model)

| | Total volatility | Temp. volatility | BA spreads | Global volume | LSE volume |
|---|---|---|---|---|---|
| Constant | -0.58 | -0.353 | 0.585 | -0.175 | -0.122 |
| | (-1.958) | (-1.076) | (1.834) | (-1.324) | (-0.662) |
| Frag. | -0.092 | 0.211 | 0.135 | 0.229 | 0.174 |
| | (-0.452) | (1.026) | (1.164) | (2.329) | (1.874) |
| Frag. sq. | 0.088 | -0.188 | -0.111 | -0.215 | -0.142 |
| | (0.463) | (-1.079) | (-1.09) | (-2.532) | (-1.766) |
| Market cap. | 0.043 | 0.014 | -0.027 | -0.006 | -0.007 |
| | (1.627) | (0.588) | (-0.861) | (-0.442) | (-0.626) |
| ME (frag.) | 0.027 | -0.043 | -0.015 | -0.06 | -0.017 |
| | (0.362) | (-0.685) | (-0.36) | (-2.138) | (-0.621) |
| $\Delta_{Frag.}$ | -0.016 | 0.048 | 0.039 | 0.043 | 0.051 |
| | (-0.277) | (0.685) | (0.978) | (1.387) | (1.732) |
| Adjusted $R^2$ | -0.061 | -0.07 | -0.037 | -0.023 | -0.022 |

*Notes*: See Table 2.3.

**Table B.10:** The effect of fragmentation on the variability of market quality for FTSE 250 firms (conditional variance model)

| | Total volatility | Temp. volatility | BA spreads | Global volume | LSE volume |
|---|---|---|---|---|---|
| Constant | -0.021 | -0.381 | 0.346 | 0.607 | 0.178 |
| | (-0.041) | (-0.682) | (1.485) | (1.204) | (0.53) |
| Frag. | -0.171 | -0.225 | -0.068 | -0.457 | -0.412 |
| | (-1.24) | (-1.884) | (-0.745) | (-2.165) | (-2.676) |
| Frag. sq. | 0.147 | 0.21 | 0.087 | 0.432 | 0.333 |
| | (1.168) | (1.833) | (0.926) | (2.409) | (2.475) |
| Market cap. | -0.043 | -0.047 | 0.004 | -0.081 | -0.084 |
| | (-1.31) | (-1.685) | (0.196) | (-3.734) | (-4.271) |
| ME (frag.) | -0.021 | -0.01 | 0.02 | -0.015 | -0.071 |
| | (-0.589) | (-0.378) | (0.883) | (-0.333) | (-1.787) |
| $\Delta_{Frag.}$ | -0.049 | -0.05 | 0.004 | -0.097 | -0.134 |
| | (-1.069) | (-1.453) | (0.157) | (-1.383) | (-2.526) |
| Adjusted $R^2$ | -0.009 | -0.011 | -0.06 | 0.048 | 0.055 |

*Notes*: See Table 2.4.

**Table B.11:** The effect of visible fragmentation and dark trading on the variability of market quality for FTSE 100 firms (conditional variance model)

| | Total volatility | Temp. volatility | BA spreads | Global volume | LSE volume |
|---|---|---|---|---|---|
| Constant | -0.879 | -0.36 | 0.663 | 0.01 | 0.2 |
| | (-2.133) | (-0.851) | (2.255) | (0.079) | (1.355) |
| Vis. frag. | 0.366 | -0.209 | -0.045 | 0.264 | 0.259 |
| | (2.588) | (-0.518) | (-0.474) | (3.244) | (2.709) |
| Vis. frag. sq. | -0.498 | 0.039 | 0.047 | -0.318 | -0.308 |
| | (-2.845) | (0.111) | (0.462) | (-3.699) | (-3.078) |
| Dark | -0.095 | -0.23 | -0.046 | -0.037 | -0.042 |
| | (-0.74) | (-2.136) | (-0.542) | (-0.838) | (-0.909) |
| Dark sq. | 0.252 | 0.393 | 0.038 | 0.057 | 0.109 |
| | (1.552) | (2.932) | (0.387) | (1.076) | (1.855) |
| Market cap. | 0.012 | 0.006 | 0.005 | -0.003 | 0.005 |
| | (0.41) | (0.22) | (0.237) | (-0.284) | (0.381) |
| ME (Vis. frag) | -0.143 | -0.17 | 0.004 | -0.061 | -0.056 |
| | (-2.029) | (-1.914) | (0.1) | (-2.795) | (-2.382) |
| ME (Dark) | 0.095 | 0.066 | -0.017 | 0.006 | 0.04 |
| | (2.477) | (2.069) | (-0.644) | (0.53) | (2.616) |
| $\Delta_{Vis.frag.}$ | 0.02 | -0.182 | -0.012 | 0.043 | 0.045 |
| | (0.378) | (-1.048) | (-0.331) | (1.56) | (1.403) |
| $\Delta_{Dark}$ | 0.15 | 0.152 | -0.009 | 0.019 | 0.064 |
| | (2.869) | (3.605) | (-0.305) | (1.225) | (3.137) |
| Adjusted $R^2$ | -0.049 | -0.055 | -0.022 | -0.012 | -0.003 |

*Notes*: See Table 2.3.

**Table B.12:** The effect of visible fragmentation and dark trading on the variability of market quality for FTSE 250 firms (conditional variance model)

| | Total volatility | Temp. volatility | BA spreads | Global volume | LSE volume |
|---|---|---|---|---|---|
| Constant | -0.436 | -0.045 | 0.163 | 0.294 | 0.054 |
| | (-1.004) | (-0.101) | (0.412) | (1.316) | (0.185) |
| Vis. frag. | -0.333 | -0.28 | 0.064 | -0.126 | -0.145 |
| | (-2.897) | (-1.97) | (0.668) | (-1.457) | (-1.377) |
| Vis. frag. sq. | 0.379 | 0.318 | -0.013 | 0.153 | 0.173 |
| | (2.169) | (1.619) | (-0.107) | (1.275) | (1.192) |
| Dark | 0.046 | -0.021 | -0.139 | -0.183 | -0.283 |
| | (0.328) | (-0.169) | (-1.645) | (-2.7) | (-3.58) |
| Dark sq. | 0.029 | 0.082 | 0.125 | 0.149 | 0.268 |
| | (0.238) | (0.752) | (1.527) | (2.749) | (4.031) |
| Market cap. | -0.042 | -0.02 | 0.026 | -0.053 | -0.052 |
| | (-1.359) | (-0.703) | (1.085) | (-3.301) | (-2.272) |
| ME (Vis. frag) | -0.093 | -0.078 | 0.056 | -0.029 | -0.036 |
| | (-1.975) | (-1.869) | (1.564) | (-1.342) | (-1.375) |
| ME (Dark) | 0.076 | 0.062 | -0.013 | -0.033 | -0.013 |
| | (2.241) | (2.185) | (-0.701) | (-1.77) | (-0.653) |
| $\Delta_{Vis.frag.}$ | -0.068 | -0.058 | 0.055 | -0.019 | -0.024 |
| | (-1.372) | (-1.431) | (1.608) | (-0.946) | (-1.032) |
| $\Delta_{Dark}$ | 0.075 | 0.061 | -0.014 | -0.034 | -0.015 |
| | (2.203) | (2.135) | (-0.757) | (-1.809) | (-0.755) |
| Adjusted $R^2$ | -0.011 | -0.02 | -0.044 | 0.04 | 0.015 |

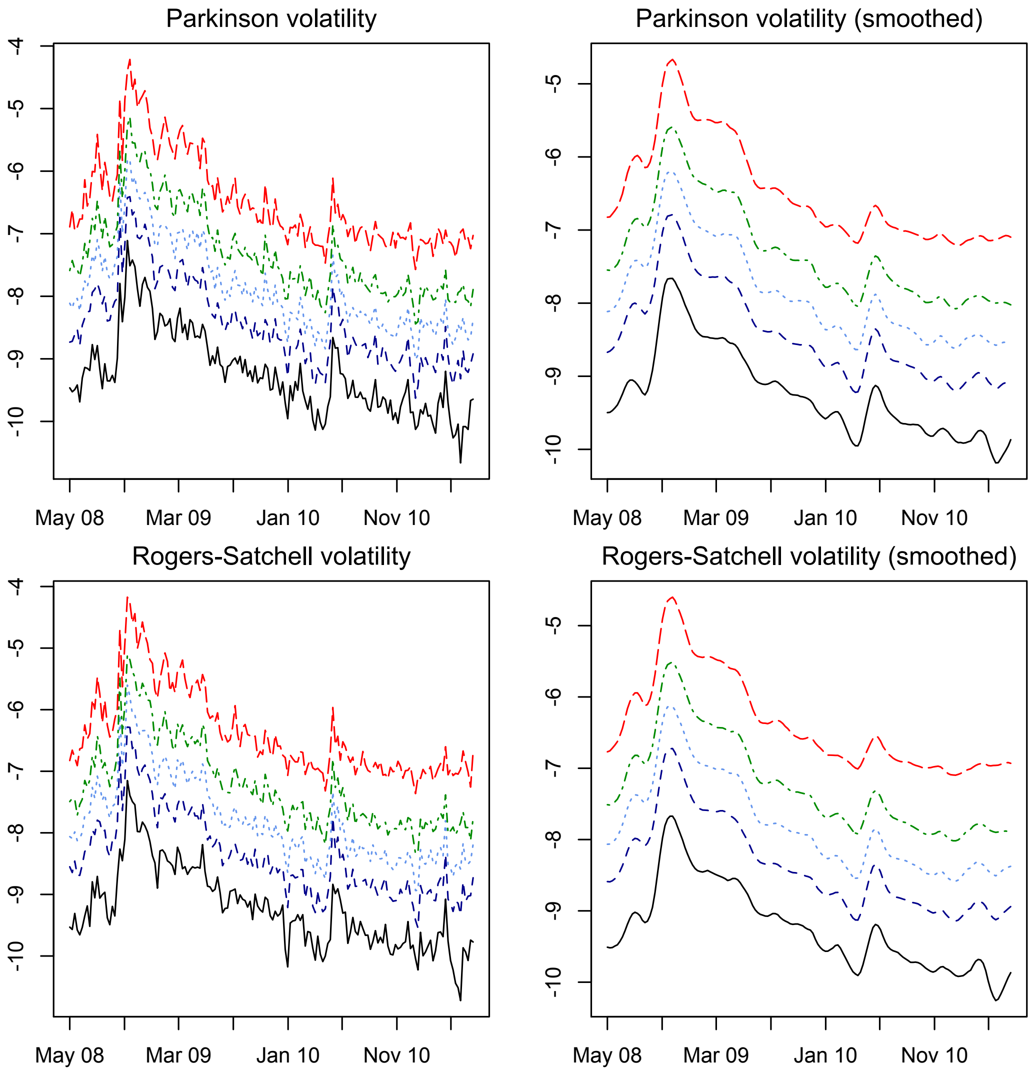
*Notes*: See Table 2.4.

**Table B.13:** The effect of fragmentation on market quality when common factor are omitted

| | Total volatility | Temp. volatility | BA spreads | Global volume | LSE volume |
|---|---|---|---|---|---|
| Constant | 4.678 | 2.375 | 0.01 | 4.619 | 4.932 |
| | (9.282) | (11.593) | (0.03) | (15.379) | (15.781) |
| Frag. | 2.803 | -0.179 | 0.98 | 0.176 | 0.741 |
| | (4.749) | (-0.541) | (3.572) | (0.528) | (2.226) |
| Frag. sq. | -3.896 | 0.25 | -1.235 | -0.055 | -2.22 |
| | (-7.488) | (0.887) | (-4.929) | (-0.179) | (-7.246) |
| Market cap. | -1.737 | -0.308 | -0.901 | -0.176 | -0.242 |
| | (-27.077) | (-14.912) | (-20.027) | (-4.541) | (-5.87) |
| ME (frag.) | -1.624 | 0.105 | -0.424 | 0.113 | -1.782 |
| | (-13.806) | (1.677) | (-6.188) | (1.19) | (-18.874) |
| $\Delta_{Frag.}$ | -0.448 | 0.03 | -0.051 | 0.129 | -1.111 |
| | (-2.409) | (0.275) | (-0.584) | (1.192) | (-10.003) |
| Adjusted $R^2$ | 0.625 | 0.015 | 0.736 | 0.681 | 0.648 |
| CSD | 0.065 | 0.051 | 0.018 | 0.149 | 0.154 |

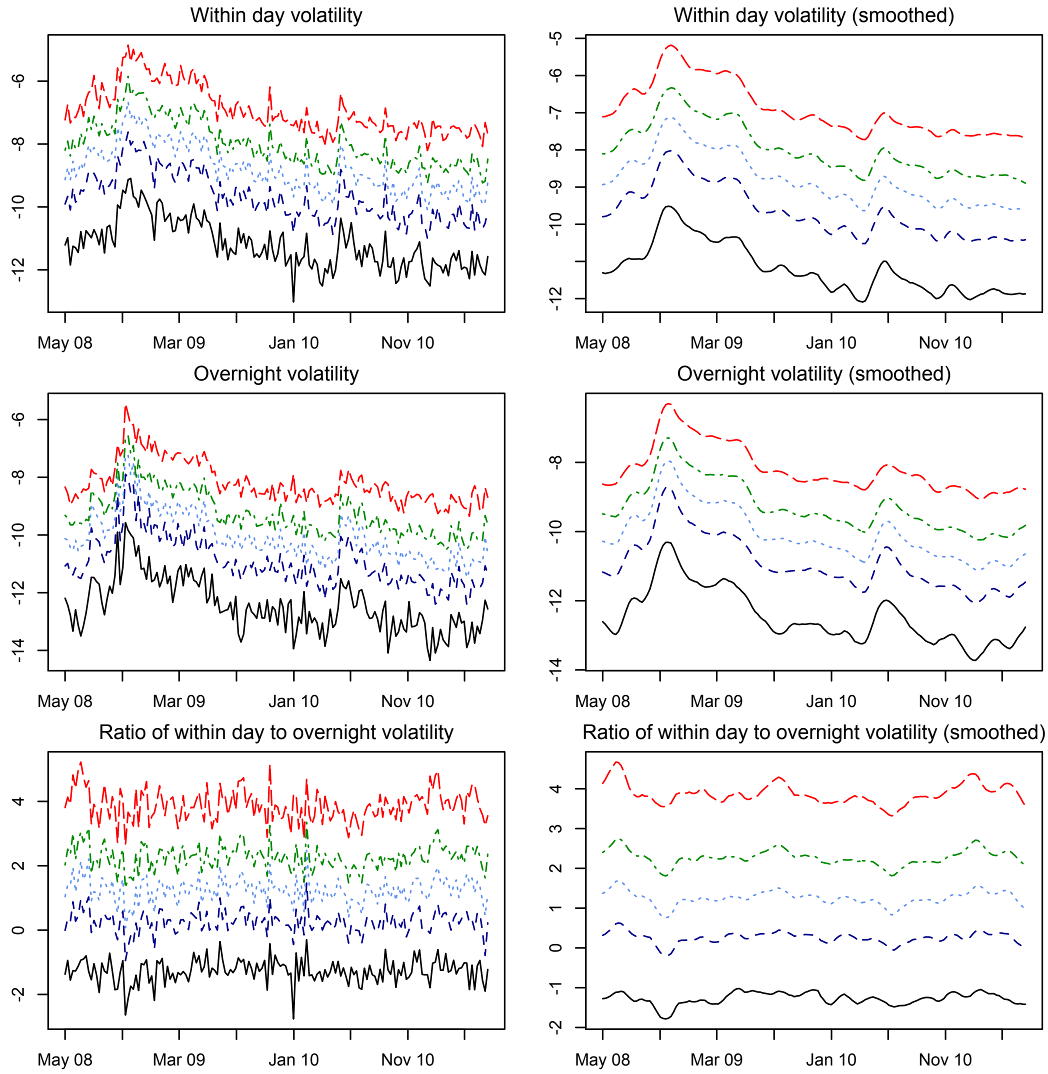
*Notes*: See Table 2.1.

**Figure B.1:** Cross-sectional quantiles for Parkinson and Rogers-Satchell volatility



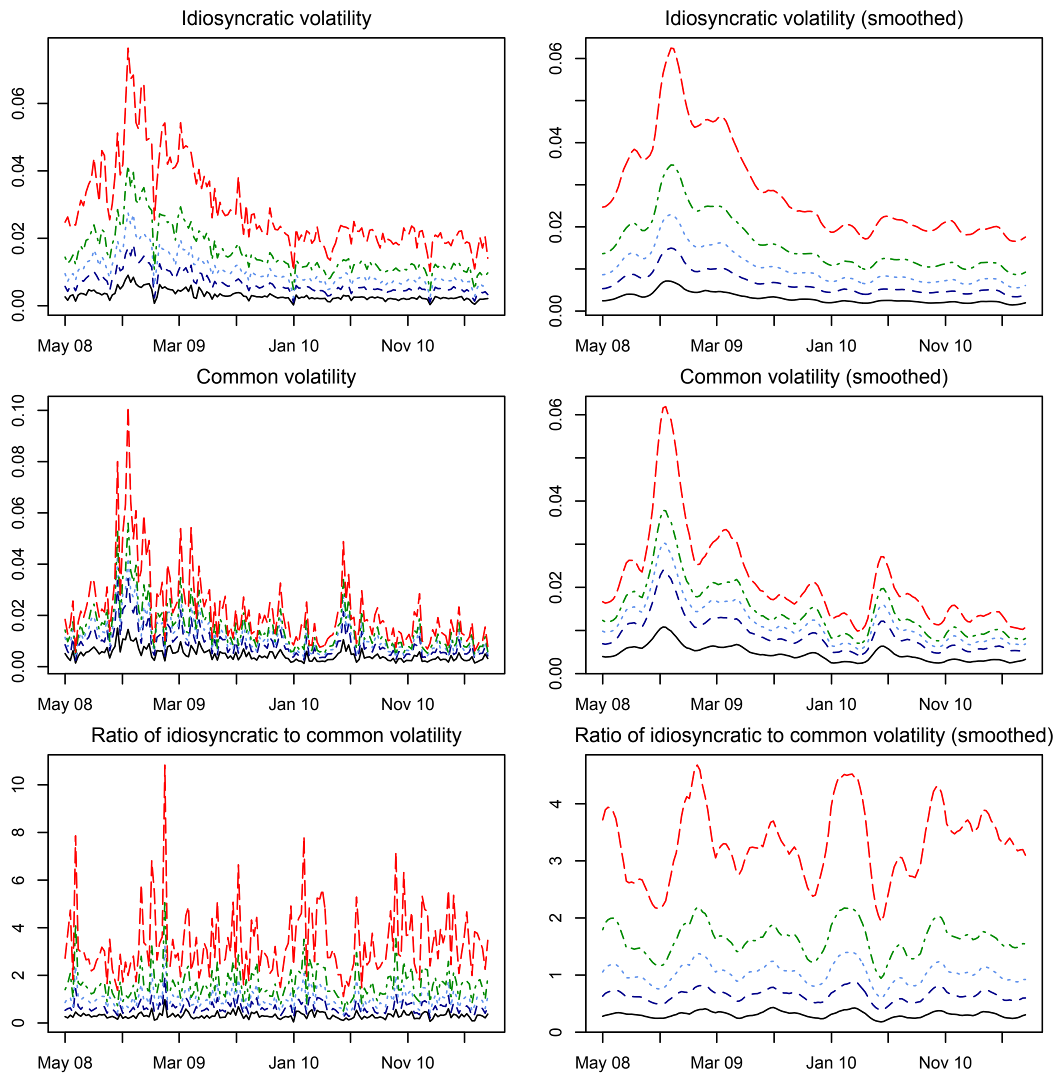

Notes: 5th, 25th, 50th, 75th and 95th percentiles of the cross-sectional distribution are shown. Volatilities are in logs. The panels on the right hand side show a nonparametric trend $m_i(t/T)$ with bandwidth parameter 0.03.

**Figure B.2:** Cross-sectional quantiles for within day and overnight volatility



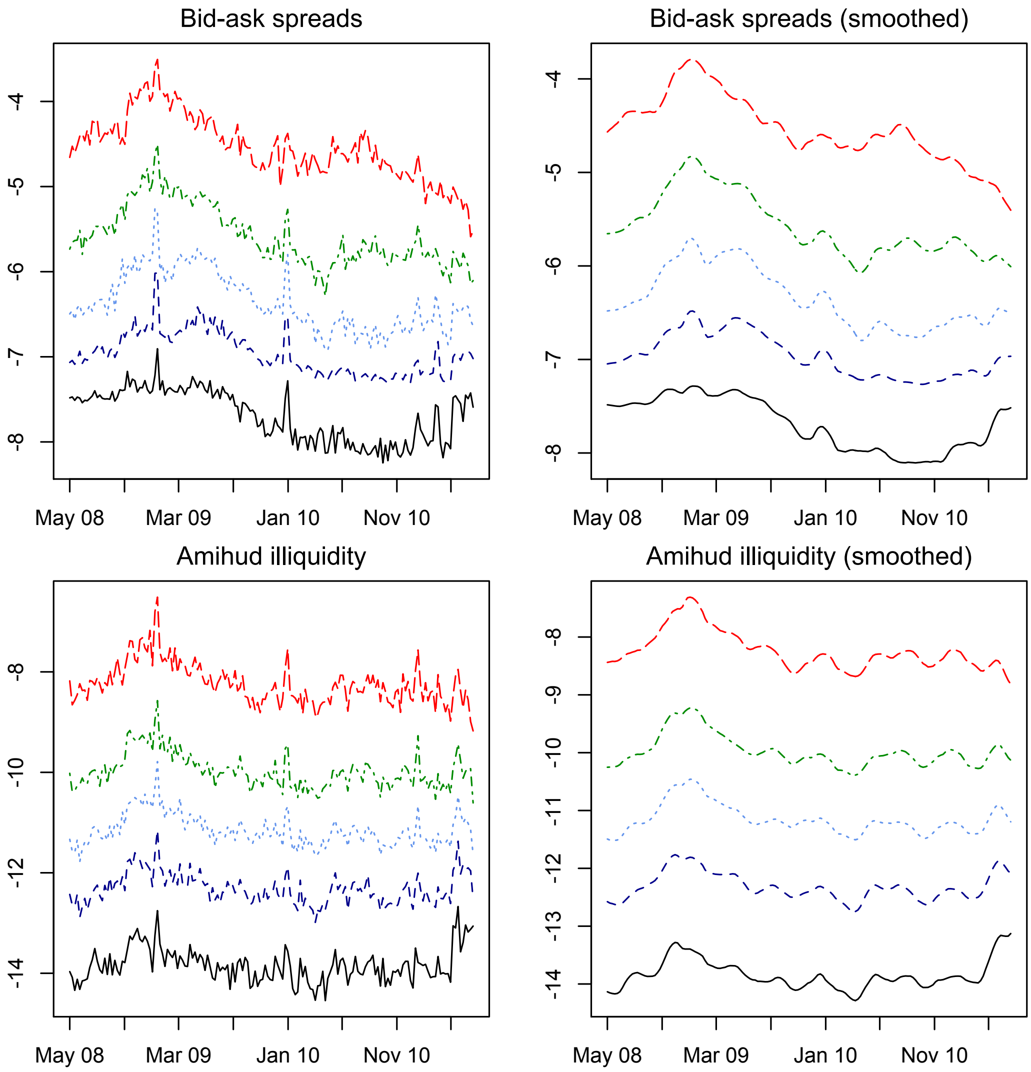

Notes: 5th, 25th, 50th, 75th and 95th percentiles of the cross-sectional distribution are shown. Within day and overnight volatilities are in logs and the ratio is the difference between the two logged variables. The panels on the right hand side show a nonparametric trend $m_i(t/T)$ with bandwidth parameter 0.03.

**Figure B.3:** Cross-sectional quantiles for idiosyncratic and common volatility



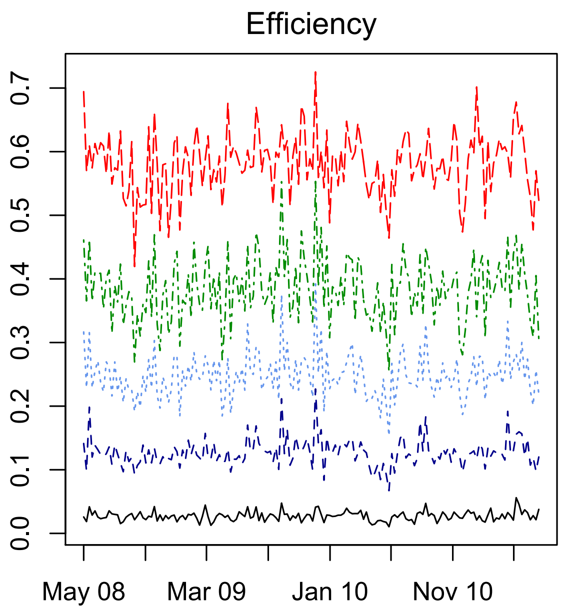

Notes: 5th, 25th, 50th, 75th and 95th percentiles of the cross-sectional distribution are shown. We took square roots of idiosyncratic and common volatilities. The panels on the right hand side show a nonparametric trend $m_i(t/T)$ with bandwidth parameter 0.03.

**Figure B.4:** Cross-sectional quantiles for illiquidity measures




Notes: 5th, 25th, 50th, 75th and 95th percentiles of the cross-sectional distribution are shown. Bid-ask spreads and Amihud illiquidity are in logs. The panels on the right hand side show a nonparametric trend $m_i(t/T)$ with bandwidth parameter 0.03.

**Figure B.5:** Cross-sectional quantiles for market efficiency measures




Notes: 5th, 25th, 50th, 75th and 95th percentiles of the cross-sectional distribution are shown. Efficiency is defined as weekly autocorrelations computed from daily data a small sample correction as in Campbell et al. (2012).

# Appendix C

# A Discrete Choice Model For Large Heterogeneous Panels with Interactive Fixed Effects

## C.1 Proofs

### Proof of Lemma 3.5.1

To derive an upper bound on the uniform rate of $\widehat{h}_t - h_{0t}$, I start by decomposing $\widehat{h}_t - h_{0t}$ into two terms that can be analyzed separately:

$$\begin{aligned}
\max_{1\leq t\leq T} \|\widehat{h}_t - h_{0t}\| &= \max_{1\leq t\leq T} \|\overline{u}_t - (A_0 - \overline{A})' d_t - (K_0 - \overline{K})' f_t\| \\
&= \max_{1\leq t\leq T} \|\overline{u}_t - (D_0 - \overline{D})' g_t\| \\
&\leq \max_{1\leq t\leq T} \|\overline{u}_t\| + \max_{1\leq t\leq T} \|(D_0 - \overline{D})' g_t\| \\
&\leq \max_{1\leq t\leq T} \|\overline{u}_t\| + \|D_0 - \overline{D}\| \max_{1\leq t\leq T} \|g_t\| \\
&\equiv T_1 + T_2
\end{aligned}$$

where $g_t = (d_t', f_t')'$, $D_0 = (A_0', K_0')'$ and $\overline{D} = (\overline{A}', \overline{K}')'$. Recall that the disturbances $u_{js}$ are IID random variables $\forall i, j, t, s$ which is required for Bernstein's inequality.
**Term $\mathbf{T_1}$** I decompose $u_{it}^j$ into $u_{it}^j = u_{it}^{j+} + u_{it}^{j-}$ where

$$\begin{aligned}
u_{it}^{j+} &= u_{it}^j I(|u_{it}^j| \leq L(N)) - E(u_{it}^j I(|u_{it}^j| \leq L(N)) \\
u_{it}^{j-} &= u_{it}^j I(|u_{it}^j| > L(N)) - E(u_{it}^j I(|u_{it}^j| > L(N))
\end{aligned}$$

Below, the terms $u_{it}^{j+}$ and $u_{it}^{j+}$ are analyzed separately.

– By Bernstein's exponential inequality applied to each element $j$ in $u_{it}^{j+}$,

$$\begin{aligned}
\Pr\left(\max_{1\leq t\leq T}\left|\sum_{i=1}^N u_{it}^{j+}\right| > \frac{\sigma\sqrt{N}}{\log T}\right) &\leq \sum_{t=1}^T \Pr\left(\left|\sum_{i=1}^N u_{it}^{j+}\right| > \frac{\sigma\sqrt{N}}{\log T}\right) \\
&\leq 2T \exp\left(-\frac{\frac{1}{(\log T)^2}}{2(1 + \frac{L(N)}{3\sigma\sqrt{N}\log T})}\right) \\
&= o(1) \qquad \text{(C.1)}
\end{aligned}$$

where the last equality follows from $\frac{L(N)}{\sqrt{N}\log T} \to 0$ under the condition that $L(N) = N^{1/3}$.

– The probability that the event $u_{it}^{j-}$ occurs can be bounded by Markov's inequality. For any $k \geq 6$ (assumption (B1)),

$$\begin{aligned}
\Pr(\max_{1\leq i\leq N}\max_{1\leq t\leq T}|u_{it}^j| > L(N)) &\leq \sum_{i=1}^N\sum_{t=1}^T \Pr(|u_{it}^j| > L(N)) \\
&\leq NT\Pr(|u_{it}^j| > L(N)) \\
&\leq NT\frac{E(|u_{it}^j|^k)}{L(N)^k}
\end{aligned}$$

$$
\begin{aligned}
&\leq NT\frac{C_u^k}{L(N)^k} \\
&= O\left(\frac{NT}{L(N)^k}\right) \\
&= o(1) \qquad \text{(C.2)}
\end{aligned}
$$

where the last equality follows from assumption (B3) and the condition that $L(N) = N^{1/3}$. This illustrates the choices for $L(N)$ and the number of moments $k$ are not independent: One the one hand, $L(N)$ has to be sufficiently small to ensure uniform consistency of $u_{it}^{j+}$. On the other hand, a large value of $L(N)$ requires a large $k$ for (C.2) to hold.

Taken together, the results in (C.1) and (C.2) imply that

$$
\max_{1\leq t\leq T}\left|\overline{u}_{it}^j\right| = O_p\left(\frac{\log T}{\sqrt{N}}\right), \forall j.
$$

**Term $\mathbf{T_2}$** By Markov's inequality,

$$
\Pr(\|\overline{D} - D_0\| > \epsilon) \leq \frac{E\|\overline{D} - D_0\|^2}{\epsilon^2}
$$

where

$$
\begin{aligned}
E\|\overline{D} - D_0\|^2 &\leq E\left\|\frac{1}{N}\sum_{i=1}^{N}(D_i - D_0)\right\|^2 \\
&= \frac{1}{N^2}\sum_{i=1}^{N}E\|D_i - D_0\|^2 \\
&= O\left(\frac{1}{N}\right).
\end{aligned}
$$

Therefore,

$$
\|D_0 - \overline{D}\| \max_{1\leq t\leq T}\|g_t\| = O_p\left(\frac{1}{\sqrt{N}}\right)
$$

because $\max_{1\leq t\leq T}\|g_t\|$ is bounded by assumption (A2). By combining the rates of the terms $T_1$ and $T_2$ it follows that $\|\widehat{h}_t - h_{0t}\|_{\mathcal{H}} = O_p\left(\frac{\log T}{\sqrt{N}}\right)$. $\square$

## Proof of Theorem 3.5.1

Because the infeasible estimator $\widetilde{\theta}_i$ is consistent (assumption (C5)), it suffices to show that estimating the unobserved factors does not affect the criterion function. We have

$$
\begin{aligned}
&\Pr\left(\sup_{\theta\in\Theta}\left|Q_T^i(\theta,\widehat{h}) - Q_T^i(\theta,h_0)\right| \geq \epsilon\right) \\
&\leq \Pr\left(\sup_{||h-h_0||\leq\delta_T}\sup_{\theta\in\Theta}\left|Q_T^i(\theta,h) - Q_T^i(\theta,h_0)\right| \geq \epsilon\right) + \Pr\left(||\widehat{h}-h_0|| > \delta_T\right) && \text{(C.3)}\\
&\to 0, && \text{(C.4)}
\end{aligned}
$$

where we have used assumption (C3) and (C6). □

## Proof of Theorem 3.5.2

To show that $\widehat{\theta}_i$ is asymptotically normal, it suffices to show that estimating the unobserved factors does not affect the limiting distribution, that is,

$$
\sup_{\theta\in\Theta}\left\|\frac{\partial^2 Q_T^i(\theta,\widehat{h})}{\partial\theta\partial\theta^\intercal} - \frac{\partial^2 Q_T^i(\theta,h_0)}{\partial\theta\partial\theta^\intercal}\right\| = o_p(1) \tag{C.5}
$$

$$
\left\|\frac{\partial Q_T^i(\theta_{i0},\widehat{h})}{\partial\theta} - \frac{\partial Q_T^i(\theta_{i0},h_0)}{\partial\theta}\right\| = o_p(T^{-1/2}) \tag{C.6}
$$

We start with equation (C.5):

$$
\begin{aligned}
&\Pr\left(\sup_{\theta\in\Theta}\left\|\frac{\partial^2 Q_T^i(\theta,\widehat{h})}{\partial\theta\partial\theta^\intercal} - \frac{\partial^2 Q_T^i(\theta,h_0)}{\partial\theta\partial\theta^\intercal}\right\| \geq \epsilon\right)\\
&\leq \Pr\left(\sup_{||h-h_0||\leq\delta_T}\sup_{\theta\in\Theta}\left\|\frac{\partial^2 Q_T^i(\theta,h)}{\partial\theta\partial\theta^\intercal} - \frac{\partial^2 Q_T^i(\theta,h_0)}{\partial\theta\partial\theta^\intercal}\right\| \geq \epsilon\right)\\
&+ \Pr\left(||\widehat{h}-h_0|| > \delta_T\right)\\
&\to 0, && \text{(C.7)}
\end{aligned}
$$

where we have used uniform consistency of $\widehat{h}$ and assumption (D2). This implies that

$$
\left\|\frac{\partial^2 Q_T^i(\theta,\widehat{h})}{\partial\theta\partial\theta^\intercal} - \frac{\partial^2 Q_T^i(\theta,h_0)}{\partial\theta\partial\theta^\intercal}\right\| = o_p(1) \tag{C.8}
$$

To show that equation (C.6) holds, consider the expansion

$$
\begin{aligned}
\frac{\partial Q_T^i(\theta_{i0},\widehat{h})}{\partial\theta} - \frac{\partial Q_T^i(\theta_{i0},h_0)}{\partial\theta} &= \frac{\partial^2 Q_T^i(\theta_{i0},h_0)}{\partial\theta\partial h^\intercal}\cdot\left(\widehat{h}-h_0\right) + R_T^i.\\
&\equiv T_1 + T_2. && \text{(C.9)}
\end{aligned}
$$

We examine the terms $T_1$ and $T_2$ separately:

**Term $\mathbf{T_1}$** Term $T_1$ can be expressed as

$$\frac{\partial^2 Q_T^i(\theta_{i0}, h_0)}{\partial\theta\partial h^\intercal} \cdot \left(\widehat{h} - h_0\right) = \frac{1}{T}\sum_{t=1}^{T} H_{2it}(\theta_{i0}, h_0)\left(\widehat{h}_t - h_{0t}\right)$$

for some weighting matrix $H_{2it} \equiv H_{2it}(\theta_{i0}, h_0)$ that is of dimensions $K_d + 2K_x \times K_x$. The rate can be obtained by Markov's inequality:

$$\Pr\left(\left\|\frac{1}{T}\sum_{t=1}^{T} H_{2it}\left(\widehat{h}_t - h_{0t}\right)\right\| > \epsilon\right) \leq \frac{E\left\|\frac{1}{T}\sum_{t=1}^{T} H_{2it}\left(\widehat{h}_t - h_{0t}\right)\right\|^2}{\epsilon^2}$$

where

$$\begin{aligned}
E\left\|\frac{1}{T}\sum_{t=1}^{T} H_{2it}\left(\widehat{h}_t - h_{0t}\right)\right\|^2 &= E\left\|\frac{1}{T}\sum_{t=1}^{T} H_{2it}\left(\overline{u}_t - (D_0 - \overline{D})^\intercal g_t\right)\right\|^2 \\
&\leq E\left\|\frac{1}{T}\sum_{t=1}^{T} H_{2it}\overline{u}_t\right\|^2 + E\left\|\frac{1}{T}\sum_{t=1}^{T} H_{2it}(D_0 - \overline{D})^\intercal g_t\right\|^2 \\
&\leq E\left\|\frac{1}{T}\sum_{t=1}^{T} H_{2it}\right\|^2\left\|\frac{1}{T}\sum_{t=1}^{T}\overline{u}_t\right\|^2 \\
&\quad + E\left\|\frac{1}{T}\sum_{t=1}^{T} H_{2it}\right\|^2\left\|\frac{1}{T}\sum_{t=1}^{T}(D_0 - \overline{D})^\intercal g_t\right\|^2 \\
&\leq O(1)O(1/(NT)) + O(1)O(1/(NT)) \\
&= O(1/(NT)), \qquad \text{(C.10)}
\end{aligned}$$

where we have used assumption (A2), (D3) and that $u_{it}$ are IID random variables with a finite variance.

Therefore,

$$\left\|\frac{1}{T}\sum_{t=1}^{T} H_{2it}\left(\widehat{h}_t - h_{0t}\right)\right\| \leq O_p(1/\sqrt{T}) \qquad \text{(C.11)}$$

**Term $\mathbf{T_2}$** The remainder term $R_T^i$ can be expressed as (with probability tending to one)

$$\begin{aligned}
R_T^i &= \left(\widehat{h} - h_0\right)^\intercal \frac{\partial^3 Q_T^i(\theta_{i0}, \overline{h})}{\partial\theta\partial h^\intercal\partial h} \cdot \left(\widehat{h} - h_0\right) \\
&= \left(\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\widehat{h}_s - h_{0s}\right)^\intercal W_{jts}(\theta_{i0}, \overline{h})\left(\widehat{h}_t - h_{0t}\right)\right)_{j=1}^{p} \\
&\leq \left(\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\sup_{\|h-h_0\|\leq\delta_T}\|W_{jts}(\theta_{i0}, h)\| \times \left\|\widehat{h}_s - h_{0s}^\intercal\right\|\left\|\widehat{h}_t - h_{0t}\right\|\right)_{j=1}^{p}
\end{aligned}$$

$$\leq T \times \left( \|\widehat{h} - h_0\|_{\mathcal{H}} \right)^2 = \frac{T(\log T)^2}{N} \tag{C.12}$$

for some intermediate values and weighting matrices $W_{jts}$ and $p = K_d + 2K_x$. $\square$

## Proof of Theorem 3.5.3

By similar arguments as in the proof of theorem 3.5.1

$$\begin{aligned}
&\Pr\left( \sup_{\theta\in\Theta} \left| \frac{1}{N}\sum_{i=1}^{N} Q_T^i(\theta,\widehat{h}) - \frac{1}{N}\sum_{i=1}^{N} Q_T^i(\theta,h_0) \right| \geq \epsilon \right) \\
&\leq \Pr\left( \sup_{||h-h_0||\leq\delta_T} \sup_{\theta\in\Theta} \left| \frac{1}{N}\sum_{i=1}^{N} Q_T^i(\theta,h) - \frac{1}{N}\sum_{i=1}^{N} Q_T^i(\theta,h_0) \right| \geq \epsilon \right) + \Pr\left( ||\widehat{h} - h_0|| > \delta_T \right) \\
&\leq \Pr\left( \max_{1\leq i\leq N} \sup_{||h-h_0||\leq\delta_T} \sup_{\theta\in\Theta} \left| Q_T^i(\theta,h) - Q_T^i(\theta,h_0) \right| \geq \epsilon \right) + \Pr\left( ||\widehat{h} - h_0|| > \delta_T \right) \\
&\to 0,
\end{aligned} \tag{C.13}$$

$\square$

## Proof of Theorem 3.5.4

The asymptotic normality proof of $\widehat{\theta}$ is similar to that of theorem 3.5.2. Asymptotic normality of $\widehat{\theta}$ follows from:

$$\sup_{\theta\in\Theta} \left\| \frac{1}{N}\sum_{i=1}^{N} \frac{\partial^2 Q_T^i(\theta,\widehat{h})}{\partial\theta\partial\theta^\intercal} - \frac{1}{N}\sum_{i=1}^{N} \frac{\partial^2 Q_T^i(\theta,h_0)}{\partial\theta\partial\theta^\intercal} \right\| = o_p(1) \tag{C.14}$$

$$\left\| \frac{1}{N}\sum_{i=1}^{N} \frac{\partial Q_T^i(\theta_{i0},\widehat{h})}{\partial\theta} - \frac{1}{N}\sum_{i=1}^{N} \frac{\partial Q_T^i(\theta_{i0},h_0)}{\partial\theta} \right\| = o_p(N^{-1/2}) \tag{C.15}$$

where equation (C.15) is implied by equation (C.10) and equation (C.14) holds because

$$\begin{aligned}
&\Pr\left( \sup_{\theta\in\Theta} \left\| \frac{1}{N}\sum_{i=1}^{N} \left( \frac{\partial^2 Q_T^i(\theta,\widehat{h})}{\partial\theta\partial\theta^\intercal} - \frac{\partial^2 Q_T^i(\theta,h_0)}{\partial\theta\partial\theta^\intercal} \right) \right\| \geq \epsilon \right) \\
&\leq \Pr\left( \sup_{\theta\in\Theta} \frac{1}{N}\sum_{i=1}^{N} \left\| \frac{\partial^2 Q_T^i(\theta,\widehat{h})}{\partial\theta\partial\theta^\intercal} - \frac{\partial^2 Q_T^i(\theta,h_0)}{\partial\theta\partial\theta^\intercal} \right\| \geq \epsilon \right) \\
&\leq \Pr\left( \max_{1\leq i\leq N} \sup_{||h-h_0||\leq\delta_T} \sup_{\theta\in\Theta} \left\| \frac{\partial^2 Q_T^i(\theta,h)}{\partial\theta\partial\theta^\intercal} - \frac{\partial^2 Q_T^i(\theta,h_0)}{\partial\theta\partial\theta^\intercal} \right\| \geq \epsilon \right) \\
&+ \Pr\left( ||\widehat{h} - h_0|| > \delta_T \right) \\
&\to 0,
\end{aligned} \tag{C.16}$$

Finally, to show that the asymptotic variance of the mean group estimator is given by the variance of the random coefficients $\Sigma_\eta$, observe that

$$\begin{aligned}
\widehat{\beta} - \beta_0 &= \frac{1}{N}\sum_{i=1}^{N}(\widehat{\beta}_i - \beta_{0i}) + \frac{1}{N}\sum_{i=1}^{N}(\beta_{0i} - \beta_0) \\
&= \frac{1}{N}\sum_{i=1}^{N}(\widehat{\beta}_i - \beta_{0i}) + \frac{1}{N}\sum_{i=1}^{N}\eta_i \\
&= o_p(1) + \frac{1}{N}\sum_{i=1}^{N}\eta_i \qquad \text{(C.17)}
\end{aligned}$$

by theorem 3.5.1. □