Essays on Financial Economics

by

Michael C. Tseng

PhD, Mathematics, The Pennsylvania State University, 2012

Dissertation Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

in the
Department of Economics
Faculty of Arts and Social Sciences

SIMON FRASER UNIVERSITY
Summer 2016

Approval

Name:	**Michael C. Tseng**
Degree:	**Doctor of Philosophy (Economics)**
Title:	***Essays on Financial Economics***
Examining Committee:	**Chair:** Simon Woodcock Associate Professor

Ramazan Gençay
Senior Supervisor
Professor _____

Robert Jones
Professor _____

Kenneth Kasa
Professor _____

Robert Grauer
Internal Examiner
Endowed University Professor
Beedie School of Business _____

Alain Chaboud
External Examiner
Principal Economist
Board of Governors of the Federal
Reserve System _____

Date Defended: 26 July 2016

Abstract

This thesis addresses three topics in modern financial economics. In econometrics, we propose a consistent estimator for a model with both smooth structural changes and abrupt structural breaks. Our methodology is particularly well-suited for modern high frequency data. In market microstructure, we show that the traditional paradigm is no longer applicable in general, in light of recent technological evolution in trading and associated change in market behavior. In financial networks, we consider determinants of systematic risk that is due to the structure and stability of the network underlying the financial system. We propose a pricing factor that captures the diversification vs. contagion risk trade-off of the interconnectedness of the network.

Keywords: Financial Economics, Econometrics

Acknowledgements

I would like to acknowledge my supervisor, Prof. Ramazan Gençay, for his support. I would also like to thank my committee members, Prof. Robert Jones and Prof. Kenneth Kasa, and Dr. Alain Chaboud of the Board of Governors of the Federal Reserve System for their generosity in sharing their time and imparting guidance.

On a personal level, I would like to thank both my family and my wife's family.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

This thesis consists of three chapters covering contemporary topics in financial economics and econometrics. An outline of each chapter is provided below.

Wavelets and High Frequency Econometrics In Chapter 2 we propose an estimation technique for a linear model with a time-varying parameter. Viewing the time-varying parameter as a functional parameter lying in an infinite-dimensional function space, our estimation techniques use properties of wavelets to achieve minimax estimation. Owing to the fact that wavelets form unconditional bases for a wide variety of function spaces, our model can accommodate serially correlated error term and arbitrary structural breaks in the time-varying parameter; it is applicable to a wide variety of situations where the economic data generating process departs abruptly from stationarity in the mean. As an empirical exercise, we apply our model to understand behavior of time-varying market risk.

A common technical thread that emerges in the structural change literature is the machinery of continuous-time stochastic processes. In order to estimate or test for a change-point, a necessary model assumption is that the sampling frequency increases with sample size. In the in-fill asymptotic framework, continuous-time processes appear in the large sample limit by applications of the Functional Central Limit Theorem. This theme extends of high-frequency econometric extends to the next chapter, where the high frequency datasets used in empirical microstructure investigations demand fitting general Itô semimartingales, rather than discrete-time time series models.

Market microstructure Financial markets is an indispensable part of the modern economy. They serve the crucial function of transferring risk across economic agents (individual investors, corporations, financial institutions, and governments) and across time and across geography. Economic growth is impossible without properly functioning financial markets. Given its macroeconomic implications, the microeconomics of the financial market as a price-formation and information aggregation mechanism has long been of great interest to

economists and regulators. This area of financial economics is known as *market microstructure* and distinguishes itself from other fields of financial economics by viewing the market through the highest resolution possible with respect to time, market structure, and market participant behavior.

The increasing pervasiveness of high speed electronic trading means that we can now observe this economics of incentive and information occurring at increasingly high—nanosecond, in some cases—frequency. This new phenomenon has given renewed impetus to re-examination of microstructure theories, which are still coming to grips with the new species of algorithmic traders whose trading is predicated on speed. The associated emergence and increasing dominance of the new breed of traders whose trading is predicated on speed –the high frequency traders (HFT's)– have fundamentally overturned traditional market microstructure paradigm. HFT now account for over half of the volume traded on stock, futures and options exchanges in the U.S. The U.S. Securities and Exchange Commission (SEC) describes HFT as "one of the most significant market structure developments in recent years" (SEC, 2010). It is a central issue of modern financial economics to better understand the economic role of HFT's and their impact on market quality and social welfare.

On the one hand, HFT seem to improve market efficiency by impounding economic information into prices at a faster rate. On the other hand, they exploit their speed advantage and engage in predatory trading, which has led to catastrophic events like the May 2010 Flash Crash –with nearly one trillion dollars disappearing from the US economy in a matter of minutes—and the Oct 15, 2014 bond market flash crash. The enhanced liquidity—as measured by traditional microstructure metrics—provided by HFT's may not be innocuous. It has been well documented that HFT's attempt to manipulate the market in a variety of ways. Some tactics used by high frequency traders for market manipulation are "stuffing", "smoking", and "spoofing". For example, "spoofing" is a bait-and-switch tactic which involves submitting and cancelling orders with no intention of execution, with the goal of swaying prices in a favorable direction. Another example, "quote stuffing", involves quickly entering and withdrawing a large number of orders in an attempt to temporarily slow down processing capacity of exchanges as well as deny access to other market participants.[1]

In Chapter 3, using approximately 500 million limit order book snapshots sampled at tick frequency we examine the interplay between market microstructure effects and the impact of algorithmic trading in the FX context. We reconcile the apparent improvement in market efficiency offered by algorithmic traders and their predatory market making activity—the first in the academic literature to do so, to the best of my knowledge.

[1] All of the above practices constitute *high frequency spam*, a mirage of specious high frequency liquidity overflow that is never intended to be executed. An immediate concern is whether these predatory strategies impede efficient price formation and damage the health of financial markets—and if so, what the appropriate regulatory response should be. The dearth of systematic knowledge on the occurrence and impact of high frequency spamming is therefore particularly alarming for regulators and exchanges. One can, witness, for example, the haphazard remedy of the SEC the day after Flash Crash and recent ad hoc microstructure rules adopted by some exchange aimed at curbing high frequency spam.

Asset pricing and financial networks While the same technological evolution in financial markets associated with the growth in electronic trading has led to a fragmentation of the financial market, on the flip side of the same coin it has also made interconnectedness between various trading platforms, market centers, and market participants an increasing important factor in market behavior. As electronic communications networks reduce search friction, increase heterogeneity of market participants. and lead to wider scope of trading, the emergence of traditional trading strategies and activities also makes network systematic and systemic risk considerations more relevant.

In Chapter 4, we extend traditional beta-pricing models to include a network pricing factor. By incorporating the network structure between assets into the portfolio decision problem, we define an asset pricing model whose components contain both a network analogue of stochastic discount factor and a corresponding pricing factor. An affine relationship is shown to hold between the network stochastic discount factor and network pricing factor. The theory thus meets a key formal benchmark. We find that exposure to the network pricing factor has two countervailing effects: diversification vs. susceptibility to undiversifiable economic shocks. Our theory predicts that the diversification effect dominates. Exposure to the network pricing factor therefore entails a risk discount, rather than premium.

Chapter 2

Wavelets and High Frequency Financial Data

2.1. Structural Change: An Brief Overview

Instability of parameters is an issue confronted by econometricians everyday. Coefficient instability reflects, for example, an economic system undergoing structural changes or an economic agent exhibiting behavioral changes (the latter as pointed out by [88]). Empirical literature has found models with constant parameters to be unsatisfactory in various contexts and devised various models to accommodate such structural changes in the underlying data generating process.

In this chapter, we consider a linear model where parameter non-constancy is restricted to a single time-varying parameter. To the best of our knowledge, theoretical estimation literature on linear models with time-varying coefficients predominantly employs predominantly kernel smoothing methods. See, for example, [29], [30], [82], and [34]. The kernel approach requires that the time-varying parameters be globally smooth over the sample period, thus ruling out abrupt structural breaks or jump behavior. While our model only admits a single time-varying parameter, the standard smoothness assumption on the time-varying parameter is dropped completely.[1] Moreover, local smoothing estimates of time-invariant parameters may show spurious time-varying behavior when applied to a misspecified model where some time-varying parameter are not globally smooth.[2] In this sense there is no trade-off between the number of time-varying parameters and type of time-varying behavior allowed between our model when compared to existing methods. When the time-varying coefficient corresponds to an intercept term, our model is a linear model with arbitrary time

[1] Formally, this means that the functional space containing the functional parameter is enlarged significantly. See Section 2.3.

[2] See Section 2.5, in particular Figures 2.11c, 2.11d, 2.11e, and 2.11f.

trend in the intercept term and overlaps with the partial linear model from [100] in the time series setting.[3]

Our approach consists of two steps. The time-invariant part of the model is estimated consistently by kernel method. With these consistent estimates in hand, the problem of estimating the time-varying parameter is converted into a nonparametric regression problem. In the second step, one has a time series of noisy observations of the time-varying parameter. The objective is then to remove the noise from the time series to reveal the true underlying trend. In our case, this trend may be an arbitrary function of time that is, for example, nonlinear and has jumps. We apply an orthogonal series estimator to this problem. In particular, we choose the orthonormal basis to be a wavelet basis and make use of their descriptive power in encoding spatial-inhomogeneity in the parameter, such as varying degree of smoothness between possible discontinuities.

Wavelets was first discovered by Daubechies in seeking an orthonormal basis for square-integrable functions that is both compactly supported and has certain smoothness properties (see [42]). The multi-resolution structure inherent in their construction makes the associated linear filters natural tool in time series analysis from a spectral perspective. A sizable body of results now testifies to the effectiveness of wavelets in this regard. For example, the unit root test of [53] and serial correlation test of [58] originate from this perspective. In contrast, the ability of wavelets to represent parsimoniously a wide variety of functions remain relatively unexplored in the econometrics. [86] introduces a test of serial correlation for covariance stationary time series based on a linear wavelet estimator of the spectral density. This nonparametric approach via wavelets allows for testing for serial correlation of arbitrary form, without smoothness assumptions on the spectral density, resulting in a powerful test. This was extended to the panel model setting by [75]. Our approach applies the same principle in the context of estimation.

In our methodology, following [44], we make use of a nonlinear wavelet estimator that is both minimax over a wide Besov scale and preserve the smoothness properties of the underlying true parameter with probability one asymptotically. The language of Besov spaces allows one to speak precisely of smooth properties more general than differentiability, or, indeed, continuity ([19]), and this language translates the ability of the wavelet estimator to faithfully extract true jump from a noisy signal containing many spurious jumps. In addition to satisfactory theoretical properties, the nonlinear wavelet estimator also offers computational advantages due to the pyramid structure of the associated filter algorithm.

We also exploit the whitening effect of wavelets on serially correlated processes in making a finite sample adjustment in extracting the time-varying trend. In this respect, the utility of wavelets is two-fold: first to de-correlate serial correlation in error terms, second to encode efficiently the underling time-varying coefficient. Both of these properties derive from the

[3]The model in [100] is a random design model while we consider a fixed design model.

time-frequency localization of wavelets, which is unique among available basis. Bases such as the classical Fourier basis or B-splines do not have this property.

There is also extensive literature on testing for structural breaks. The flexibility of wavelet estimator can complement existing testing methods by supplying supplemental information within the class of models being considered. The seminal paper [10] introduced tests for structural break in a general partial sample general method of moments framework. The Andrews test is formally designed against alternatives where the structural change occurs in a specified interval. However, information about location of change point for the parameter in question may not be available to the econonometrican. For example, the issue at hand could be general model adequacy. Or the structural break may be caused by a policy change, the lag-effect of which is is unknown. For the class of model considered in this paper, the Andrews alternative is the special case where the time-varying coefficient is a piecewise constant function of time. An estimator that captures the true jumps can be useful in providing possible locations of structure breaks in the Andrews alternative. When the time-varying coefficient is piecewise constant, our model is also a special case of the linear models with partial structural breaks considered in [13] and [95]. For finding the number of breaks, [13] provides a test that proceeds stepwise by considering, for each l, whether the model has l versus $l+1$ structural breaks. In procedures of estimating the number and then locations of the breaks, considerable computation may be circumvented by inspecting the a consistent (in the appropriate sense) estimate of the time-varying parameter as a preliminary step.

In empirical research, the parameter instability might be addressed by either introducing time series specifications for the parameter or fitting elaborate non-linear alternatives. Neither approach directly confronts the issue if the true source of misspecification is parameter instability in time. Furthermore, the econometrician may be forced to impose additional structures on the model. In such situations, a simple linear model that allows for both abrupt structural breaks and smooth structural changes, should one be available, deserves consideration before other alternatives. To cite one example, [91] rejects via the Chow test a constant relationship between real house prices and real disposable income, which can be attributed to changes to the economy during the period examined such as financial deregulation and interest rate fluctuation. Similarly, [65] found evidence of susceptibility of United States housing prices to structural changes. As a remedy, various time series specifications, such as state space or GARCH models, have been devised (see, for example, [27]). Such time series specifications introduces additional structure in the model that is not implied by rejection of the Chow test. A linear model with structure breaks is the Chow alternative.

The rest of this chapter is organized as follows. In Section 2.2, we specify the model and recall necessary facts in estimating the time-invariant part of the model. Section 2.3 summarizes relevant aspects of wavelet theory. Section 2.4 discusses our approach to the nonparametric regression problem for the time-varying parameter. Certain facts regarding

the classical nonparametric model is recast in a way that makes clear the superiority of the wavelet approach relative to other estimators. In a model with weakly dependent data, there is a wavelet estimator that preserves the smoothness of the true parameter in large sample and we derive its rate of convergence. We also consider a finite sample adjustment where a model with long range dependence decomposes asymptotically into white noise models across different scales. This suggest naturally a scale-dependent wavelet estimator. Section 2.5 contains simulation results and also demonstrates existing methods (local smoothing and spline) breakdown when parameters are not smoothly varying in time. Section 2.6 applies our method to financial data in estimating high frequency systematic risk. Appendix A contains the proofs.

2.2. Model and Estimation of Time-Invariant Parameters

In this section we state model assumptions and describe how the time-invariant part of the model can be consistently estimated by a Nadaraya-Watson local regression. We consider the linear model with time series data

$$Y_t = \sum_{i=0}^{m} \beta_i x_t + \gamma(t) x_{m+1,t} + \epsilon_t, \tag{2.1}$$

with the following model assumptions:

Assumption 1.

(i) *The parameter* $\beta = (\beta_0, \cdots, \beta_m)' \in \mathbb{R}^{m+1}$ *is time-invariant.*

(ii) $(x_{1,t}, \cdots, x_{m+1,t}, \epsilon_t)' = (X_t', \epsilon_t)'$ *is an* α*-mixing.*

(iii) $\sup_{1 \le i \le m+1} \sup_t |x_{i,t}|^{4+\delta} < \infty$ *for some* $\delta > 0$.

(iv) $\inf_t \|E[X_t X_t']\| > \eta > 0$.[4]

(v) $E[X_t \epsilon_t] = 0$ *for all* t.

(vi) $\sup_t E[|\epsilon_t|^{4+\delta'}] < \infty$ *for some* $\delta' > 0$.

(vii) *The time series* $\{(x_{1,t}, \cdots, x_{m+1,t}, Y_t)\}$ *is sampled at the rate* $\frac{1}{n}$ *on the interval* $[0,1]$, *i.e. at times* $\frac{1}{n}, \frac{2}{n}, \cdots, 1$.

(viii) *The time-varying parameter* $\gamma(t)$ *is Riemann integrable on* $[0,1]$.

Under the above assumptions, the time-invariant part of the model β can be estimated by drawing $t_0 \in (0,1)$ at random and carrying out a local kernel regression. Let $K(t)$ denote

[4]For an $n_1 \times n_2$ matrix A, $\|A\|$ denotes the operator norm $\|A\| = \sup_{v \in \mathbb{R}^{n_2},\, \|v\| \le 1} \|Av\|$. $\|A\|$ can be characterized as the square-root of largest singular value of A.

the kernel used and $K_h(t) = \frac{1}{h}K(\frac{t}{h})$ corresponding to bandwidth h. Then $\hat{\beta}_{n,h_n}$ is obtained by minimizing weighted sum of residual squares:

$$(\hat{\beta}', \hat{\gamma}(t_0))'_{n,h_n} = \underset{(b',c)'}{\operatorname{argmin}} \sum_{t=1}^{n} K_{h_n}(t - t_0)(Y_t - X_t'(b', c)')^2. \tag{2.2}$$

Standard assumptions are made regarding the kernel K:

Assumption 2.

(i) K is symmetric, has compact support, and continuous with $\int K(t)dt = 1$.

(ii) The kernel bandwidth h_n satisfies $h_n \to 0$ and $nh_n \to \infty$.

Assumptions 1 and 2 are (weaker than, in parts) standard assumptions in the time-varying parameter estimation literature. The data generating process need not be stationary. Assumption 1(ii) does not impose any size restriction on the α-mixing coefficient. Assumption 1(iv) is a uniform non-multicolinearity assumption slightly stronger than uniform positive-definiteness. The latter gives consistency in the non-weighted regression. Similarly, while $\frac{1}{n}\sum_{t=1}^{n} E[X_t\epsilon_t] \to 0$ is sufficient to achieve consistency in the non-weighted regression, contemporaneous exogeneity of regressors—Assumption 1(v)—is needed to accommodate the use of a kernel.[5] Assumption 1(viii) ensures that the time-varying parameter γ is continuous Lebesgue-almost everywhere, therefore making it suitable for Nadaraya-Watson regression. The moment restrictions Assumptions 1(iii) and 1(vi) can be relaxed slightly at the expense of Assumption 1(ii) being strengthened to ϕ-mixing. Assumption 1(vii) and 2(ii)—which amounts to that the sampling frequency within a kernel bandwidth can be made arbitrarily high as sample size becomes large—are necessary in general for consistent estimation.[6]

Theorem 2.2.1. *Under Assumptions 1 and 2, the kernel estimate $\hat{\beta}_{n,h_n}$ defined by Equation 2.2 at any t_0, randomly chosen with respect to the uniform distribution on $[0, 1]$, is a consistent estimator of β.*

2.3. Wavelets and Function Spaces

Having estimated the time invariant part of the model by local smoothing, we now estimate the time-varying parameter γ by global smoothing. Specifically, a non-linear orthogonal series estimator will be used, with wavelets as chosen basis. Unlike other families of

[5] Regressors in our model can be endogenous if instrumental variables are included as part of the data generating process. The consistency result for the time-invariant parameters, Theorem 2.2.1, remains true with minor modification of arguments. Subsequent results only rely on consistency of $\hat{\beta}_{n,h_n}$ and are not affected.

[6] See, for example, [101].

basis such as the Fourier basis or splines, wavelets allow γ to have unrestricted time-varying behavior. We summarize in this section relevant aspects of wavelet theory.

Let ψ be a Daubechies mother wavelet of compact support having r vanishing moments, r continuous derivatives and unit L^2-norm (see [42]). An orthonormal basis of the Hilbert space $L^2(\mathbb{R})$ is generated using ψ by integer translations and dyadic dilations by defining ($\mathbb{Z}$ denotes the integers):

$$\psi_{jk} = 2^{\frac{j}{2}}\psi(2^j x - k), \quad j,k \in \mathbb{Z}.$$

For a function in $L^2(\mathbb{R})$, its wavelet decomposition is $f(x) = \sum_{j,k}\theta_{jk}\psi_{jk}(x)$ where $\theta_{jk} = \int f(x)\psi_{jk}(x)dx$. For $L^2[0,1]$, an appropriate modification can be made on a subset of $\mathbb{Z}\times\mathbb{Z}$ so that the resulting $\{\psi_{jk}\}$ form an orthonormal basis of $L^2[0,1]$, which contains the time-varying parameter γ in our model (see [39]).

Wavelets give rise to a *multiresolution analysis*, a Gram-Schmidt type algorithm for the Hilbert space $L^2([0,1])$, which yields numerically fast computations involving the orthonormal basis $\{\psi_{jk}\}$. It can be shown that there exists a $\mathbb{Z}$-indexed nest of subspaces $\cdots V_1 \subset V_0 \subset V_{-1} \subset \cdots$ such that $\cap_n V_n = \{0\}$, $\overline{\cup V_n} = L^2[0,1]$, and $f(\cdot) \in V_j$ if and only if $f(2\cdot) \in V_{j-1}$. The subspace V_{j-1} is of resolution twice that of V_j.[7] The map $f(x) \mapsto \frac{1}{\sqrt{2}}f(2x)$ is an isometric embedding from V_j to V_{j-1}. There exists a function $\phi_{0,0} \in V_0$ such that $\{\phi_{0,k}(\cdot) = \phi_{0,0}(\cdot - k)\}$ is an orthonormal basis of V_0. The orthonormal basis $\{\psi_{j,k}\}_{j,k\in\mathbb{Z}}$ generated using the mother wavelet has the property that where for each j, $\{\psi_{j,k}\}_{j\in\mathbb{Z}}$ is an orthonormal basis of the orthogonal complement $V_{j-1} \ominus V_j$.[8] In our context, the multiresolution analysis of the time-varying parameter γ is used in a finite sample adjustment for serial correlation in the error term.[9]

Wavelet bases are unconditional bases for a wide variety of functional spaces which are convenient in modelling spatially inhomogeneous functions. Empirically, this means that wavelets provide parsimonious representation of functions whose smoothness can, for example, vary between possible discontinuities. We work with *Besov spaces* and refer to [90] and [107] for more details on functional analytic properties of wavelets. Unlike the classical Sobolev or Hölder spaces, Besov spaces go beyond continuity and allow one to quantify smoothness of, for instance, cádlág functions. Allowing γ to lie in a Besov space means that our model accommodates structural breaks of arbitrary type.

The *Paley-Littlewood* definition of Besov space is as follows ([90]). Let $\mathcal{S}'$ be the space of tempered distributions, i.e. the topological vector space dual of the Schwartz space $\mathcal{S}$ of C^∞-test functions ([56]). Fix $\Psi, \{\Phi_n\}_{n\geq 0} \subset \mathcal{S}$ such that their Fourier transforms $\mathcal{F}(\Psi), \{\mathcal{F}(\Phi_n)\}$ form a partition of unity subordinate to the open cover $A_0 = (-1,1)$,

[7] More generally, the scaling constant 2 can be replaced by any integer.

[8] There exists a function $\phi_{0,0} \in V_0$, called the *father wavelet*, such that its integer shifts, $\{\phi_{0,k}(\cdot) = \phi_{0,0}(\cdot - k)\}$, is an orthonormal basis of V_0. One can then write $\phi_{0,0} = \sum_{k\in\mathbb{Z}} h_k\phi_{-1,k}$. For $j = 0$, $\psi_{0,0} = \sum_k (-1)^k h_{-k+1}\phi_{-1,0}$ and $\psi_{0,k'}(\cdot) = \sum_k (-1)^k h_{-k+1}\phi_{-1,0}(\cdot - k')$. This specifies $\{\psi_{j,k}\}_{j,k\in\mathbb{Z}}$.

[9] See Section 2.4.1.

$A_n = \{2^{n-1} < |\xi| < 2^{n+1}\}$. So an element $f \in \mathcal{S}'$ can be written as $f = \Psi * f + \sum_{n\geq 0} \Phi_n * f$, with $*$ denoting convolution. f is said to lie in the inhomogeneous Besov space $B^{\alpha}_{p,q}$ if

$$\|\Psi * f\|_{L^p} + (\sum_{n\geq 0} (2^{n\alpha}\|\Phi_n * f\|_{L^p})^q)^{\frac{1}{q}} < \infty.$$

As a corollary of (sufficiently smooth) wavelets forming an unconditional basis for $B^{\eta}_{p,q}$, f lies in the inhomogeneous Besov space $B^{\eta}_{p,q}$ if

$$\sum_{j\geq j_0} \left(2^{jq(\eta+\frac{1}{2}-\frac{1}{p})}\|\theta_{jk}\|_{l^p}\right)^q < \infty. \tag{2.3}$$

Example In the model being considered, let the time-varying parameter $\gamma(t) = 1_{[0,\frac{1}{2}]}(t)$, the indicator function on $[0, \frac{1}{2}]$. This is a parameter that has an abrupt structural break at $t = \frac{1}{2}$ but constant otherwise. Such a γ might be, for instance, an alternative hypothesis of the Chow test. Let $\mathcal{F}(\gamma)$ denote the Fourier transform of γ, then

$$\mathcal{F}(\gamma)(\xi) = \frac{1}{4} e^{-2\pi i \cdot \frac{1}{4}\xi} \cdot \frac{\sin\frac{\pi\xi}{4}}{\frac{\pi\xi}{4}}.$$

Choose $p = 2$, then in the above notation, the L^2-norm of the n-th Paley-Littlewood term is, up to a multiplicative constant independent of n,

$$\|\Phi_n * g\|_{L^2} \approx 2^{-\frac{n}{2}}.$$

Therefore $\gamma \in B^{\eta}_{2,q}$ if and only if

$$\sum_n 2^{n(\eta-\frac{1}{2})q} < \infty,$$

For $q \in [1, \infty]$, this is true whenever $\eta < \frac{1}{2}$. For $\eta = \frac{1}{2}$, one must have $q = \infty$.

More generally, almost all sample paths of a Lévy process, which can be, for example, a càdlàg function where each continuous piece is a Brownian sample path, lie in a Besov space ([72]). Almost all sample paths of a Brownian motion belong to the suitable Besov spaces $B^{\eta}_{p,q}$ with $1 \leq p, q \leq \infty$, $\frac{1}{p} < \eta \leq \frac{1}{2}$. Such sample paths have the property, for example, of crossing 0 infinitely many times on the time interval $(0, \epsilon)$ for ϵ arbitrarily small. From a practical perspective, this descriptive power of Besov spaces means that, our model places no restrictions on the time-varying parameter γ, allowing for features such as smooth structural changes, abrupt jumps, or combinations thereof.

The family of Besov spaces contains both L^2-Sobolev spaces (the cases $p = q = 2$) and Hölder spaces (the case $p = q = \infty$). The assumption of twice-differentiability that is common in the time-varying coefficient literature correspond to the case $\eta = 2$ and $p = q = \infty$.

2.4. Estimation of Time-Varying Parameter

Plugging in the consistent estimates of $\beta_1, \cdots, \beta_m$ given by Theorem 2.2.1 and re-writing the model of Equation 2.1 gives, for $t = 1, \cdots, n$,

$$\begin{aligned} \frac{Y_t - \sum_i \hat{\beta}_i x_{i,t}}{x_{m+1,t}} &= \frac{Y_t - \sum_i \beta_i x_{i,t}}{x_{m+1,t}} + \frac{\sum_i (\beta_i - \hat{\beta}_i) x_{i,t}}{x_{m+1,t}} \\ &= \underbrace{\gamma(t) + \frac{\epsilon_t}{x_{m+1,t}}}_{\text{De-noising problem}} + \underbrace{\frac{\sum_i (\beta_i - \hat{\beta}_i) x_{i,t}}{x_{m+1,t}}}_{\text{Estimation error of } \beta_1, \cdots, \beta_m} . \end{aligned} \tag{2.4}$$

The above expression suggests that the problem of estimating γ is, conditional on the true time-invariant parameters, a nonparametric regression problem with additional estimation error from the plug-in first step. We first consider the nonparametric regression problem, showing the faithfulness of the wavelet estimator with respect to the smoothness of γ and establishing its minimax rate of convergence, before incorporating estimation error.

2.4.1. The Nonparametric Regression Problem

Assumption 3.

(i) The time-varying parameter γ lies in a Besov space $B^{\eta}_{p,q}[0,1]$ as defined in Equation 2.3.

(ii) The marginal distribution of the regressor x_{m+1} has support bounded away from zero.

(iii) $(x_{m+1,t}, \epsilon_t)'$ is α-mixing of size $-\frac{r}{r-2}$ for some $r > 2$, $E[\epsilon_t | x_{m+1,t}] = 0$ for all t, and $E[(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{\epsilon_t}{x_{m+1,t}})^2] \to \sigma^2$.

Assumption 3(iii) strengthens Assumption 1(ii), which places no restriction on mixing rate, and Assumption 1(v), regarding the components $(x_{m+1,t}, \epsilon_t)'$ of the data generating process $(X_t', \epsilon_t)'$. From a modelling point of view, while endogeneity of $x_{i,t}$, $1 \leq i \leq m$ is still allowed in the model, to obtain minimax estimation of γ we now insist that the regressor $x_{m+1,t}$ corresponding to the time-varying parameter be contemporaneously strictly exogenous. A sufficient condition for convergence of $E[(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{\epsilon_t}{x_{m+1,t}})^2]$ is, for example, strict stationarity of $(x_{m+1,t}, \epsilon_t)'$.

The term $\gamma(t) + \frac{\epsilon_t}{x_{m+1,t}}$ from Equation 2.4 can be viewed as noisy observations of γ. In more compact notation, we rewrite it as

$$Z_t = \gamma(t) + u_t, \tag{2.5}$$

where $u_t = \frac{\epsilon_t}{x_{m+1,t}}$. Allowing for jumps at unknown times necessarily means that one must abandon the notion of pointwise consistency. Instead, we measure the loss by the norm

$\|\cdot\|_{L^2}$ on $L^2[0,1]$. The corresponding risk is the mean integrated square error (MISE). For a given estimator $\hat{\gamma}$, the MISE is

$$\mathbb{E}[\|\hat{\gamma}-\gamma\|_2^2] = \mathbb{E}[\int_0^1 |\hat{\gamma}(t)-\gamma(t)|^2 dt].$$

The benchmark is the *minimax risk* over an infinite dimensional collection of possible γ's. The (asymptotic) minimax risk over a subset $\mathcal{F}$ (e.g. a subspace such as $B^{\alpha}_{2,q}$ or a family of subspaces) is defined by

$$\lim \inf_{n\to\infty} \inf \sup_{\gamma\in\mathcal{F}} \mathbb{E}[\|\tilde{\gamma}-\gamma\|_2^2],$$

where $\inf_{\tilde{\gamma}}$ denotes infimum over all $\mathcal{F}$-valued maps measurable with respect to data.

Mixing $\{\epsilon_t\}$

The nonparametric regression of Equation 2.5 with homoskedastic white noise u_t and the infinite dimensional Gaussian sequence model of Equation 2.6 below are shown formally to be statistically equivalent by [28]. We offer a development showing these two models have the same minimax risk, in the non-stationary setting, that explicates the utility of wavelets, first as the basis used in an orthogonal series estimator and second in facilitating the extension to $\{u_t\}$, from Equation 2.5, with long range dependence.

From nonparametric regression to filtering model: For a given n, define a stochastic process on $[0,1]$ as follows:

$$F_t^{(n)} = \frac{1}{n}\sum_{i=1}^{\lceil nt\rceil} Z_i = \frac{1}{n}\sum_{i=1}^{\lceil nt\rceil}\gamma(t) + \frac{1}{n}\sum_{i=1}^{\lceil nt\rceil} u_t.$$

Viewed as a sequence of probability measures on the space $D[0,1]$ of càdlàg functions on $[0,1]$, the drift term $\frac{1}{n}\sum_{i=1}^{\lceil nt\rceil}\gamma(t) \Rightarrow \int_0^t f(t)dt$, where $\Rightarrow$ denotes weak convergence of probability measures.[10] We recall the following Functional Central Limit Theorem ([110]):

Theorem 2.4.1. *Under Assumptions 1(vi) and 3,*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{\lceil nt\rceil} u_t \Rightarrow \sigma B_t,$$

where σ^2 is as in Assumption 3(iii) and B_t is the standard Brownian motion on $[0,1]$.

In practice, the error variance σ^2 is estimated from the highest level wavelet coefficients, the most noisy part of the noisy observations of γ.

[10] $D[0,1]$ is equipped with the Skorokhod metric and the corresponding Borel σ-algebra.

By Theorem 2.4.1, the sequence of processes $\{F_t^{(n)}\}$ is asymptotically equivalent to the sequence of the Itô processes

$$dF_t = \gamma(t)dt + \sigma\frac{1}{\sqrt{n}}dB_t,\ t \in [0,1].$$

From filtering model to Gaussian sequence model: We make use of the Lévy-Ciesielski-Itô construction ([87], [38], [79]) of the Brownian motion B_t.[11] Let $\{\psi_j\}$ of $L^2[0,1]$ be an arbitrary orthonormal basis, then the standard Brownian motion on $[0,1]$ can be expressed as

$$dB_t = \sum_j w_j \psi_j(t)dt$$

where $\{w_j\}$ is standard Gaussian white noise and the series converges in the mean square sense.[12] Therefore

$$\int_0^1 \psi_j(t)dF_t = \int_0^1 \psi_j(t)\gamma(t)dt + \int_0^1 \psi_j(t)\cdot\sigma\frac{1}{\sqrt{n}}dB_t$$

where $\int_0^1 \psi_j(t)\cdot\sigma\frac{1}{\sqrt{n}}dB_t = \sigma\frac{1}{\sqrt{n}}w_j$ by orthonormality of $\{\psi_j\}$. Therefore, in this stochastic sense, the Fourier transform of the filtering model taken with respect to $\{\psi_j\}$ is the Gaussian sequence model

$$X_j = \theta_j + e_j, \tag{2.6}$$

where $(\theta_j) \in l^2(\mathbb{N})$, $e_j \sim \mathcal{N}(0, \frac{\sigma^2}{n})$, $j = 1, 2, \cdots$.[13]

We summarize the above in a theorem:

Theorem 2.4.2. *Under Assumptions 1(vi), 1(vii) and 3, the following three (sequences of) models have the same asymptotic minimax risk:*

(i)

$$Z_t = \gamma(t) + u_t$$

where Z_t is sampled at the rate $\frac{1}{n}$ for a given n.

(ii) The filtering model where the drift of the Itô process

$$dF_t = \gamma(t)dt + \sigma\frac{1}{\sqrt{n}}dB_t$$

is to be estimated, sampled at the rate $\frac{1}{n}$ for a given n.

[11] We recall the details of this construction in the Appendix.
[12] Strictly speaking, we take a modification of the process $\sum_j w_j\psi_j(t)dt$ that has continuous sample paths.
[13] $l^2(\mathbb{N})$ denotes the Hilbert space of square summable sequences.

(iii) The Gaussian sequence model

$$X_j = \theta_j + e_j,$$

where $(\theta_j = \int_0^1 \gamma\psi_j) \in l^2(\mathbb{N})$ *for any orthonormal basis* $\{\psi_j\}$, $e_j \sim \mathcal{N}(0, \frac{\sigma^2}{n})$, $j = 1, 2, \cdots$.

Remark 2.4.3. *The above development, with passage through the intermediate filtering model, generalizes to the case of long range dependent* $\{\epsilon_t\}$. *In the mixing case, model equivalence holds for any basis, as stated in Theorem 2.4.2(iii), with the means of the resulting Gaussian sequence model being the expansion coefficients with respect the chosen basis. The effectiveness of wavelet lies in that a wide variety of functions can be parsimoniously encoded by their wavelet coefficients. The long range dependent case, however, the additional property of wavelets to de-correlate serial dependence is required to arrive at a similar Gaussian sequence model.*

The sequence (θ_j) are the expansion coefficients of γ with respect to the chosen basis (ψ_j) whose empirical counterpart is (X_j). The seminal estimation result in this setting is due to Pinsker, who used the classical Fourier basis and showed that minimax risk can be achieved over Sobolev ellipsoids by shrinking the empirical Fourier coefficients ([96]). While a shrinkage estimator is natural in this setting, it is impossible to extend the minimax result beyond Sobolev spaces.[14] Wavelets, however, allows one to extend beyond continuous functions. Define the *soft threshold function* η_λ with threshold λ by

$$\eta_\lambda(x) = \text{sgn}(x)(|x| - \lambda)_+.$$

The *wavelet universal threshold estimator* $\hat{\gamma}_n$, where n is sample size, applies $\eta_\lambda(\cdot)$ to each empirical wavelet coefficient with threshold $\lambda = \hat{\sigma}\sqrt{\frac{2\log n}{n}}$, where $\hat{\sigma}$ is the median absolute deviation estimate of σ from the highest level of wavelet coefficients. It was shown in [44] that not only does wavelet universal threshold estimator achieve minimax risk over a Besov scale (up to a log factor), the estimated function $\hat{\gamma}$ is as smooth as γ with probability approaching 1.[15] The minimax property in turn passes through the equivalence of Theorem 2.4.2:

Theorem 2.4.4. *Suppose Assumption 1(vi), 1(vii) and 3 hold. Let* ψ, *the mother wavelet that generates the wavelet basis used, have* r *vanishing moments and* r *continuous derivatives, where* $r > \max\{1, \eta\}$. *Let* $R_n(p, q, \eta, L)$ *denote the minimax risk over the Besov ball* $B^\eta_{p,q}(L)$. *Then the wavelet universal threshold estimator* $\hat{\gamma}_n$ *for the nonparametric regression model has the following properties:*

[14] We have the following fact from Fourier analysis: For any $g \in L^2[0, 1]$, there exists a continuous h on $[0, 1]$ such that all the Fourier coefficients of h are larger than those of g ([81]). Therefore shrinking the Fourier coefficients does not in general preserve smoothness.

[15] Strictly speaking, the statements from [44] applies to the sequence of truncated finite dimensional Gaussian sequence models. The gap with our formulation is bridged with Lemma A.1 in the Appendix.

(i)

$$\frac{\sup_{\gamma \in B^{\eta}_{p,q}(L)} \mathbb{E}[\|\hat{\gamma}_n - \gamma\|^2]}{R_n(p,q,\eta,L)} = O(2\log n + 1)$$

for all $1 \leq p, q \leq \infty$, $0 < L < \infty$, *and* $\eta_0 < \eta < r$ *where*

$$\eta_0 = \max\{\frac{1}{p}, 2(\frac{1}{p} - \frac{1}{2})_+\}.$$

(ii) There exists a constant C *such that*

$$Prob\{\|\hat{\gamma}_n\|_{B^{\eta}_{p,q}} \leq C\|\gamma\|_{B^{\eta}_{p,q}}\} \to 1.$$

Even as one does away with the notion of pointwise estimation, Theorem 2.4.4(ii) guarantees that the estimate $\hat{\gamma}_n$ is as smooth as the true γ, as measured by the Besov norm $\|\cdot\|_{B^{\eta}_{p,q}}$, with probability approaching 1. Empirically speaking, there are no spurious jumps in the estimate.

Long range dependent $\{u_t\}$

The model equivalence described in Theorem 2.4.2 and the corresponding wavelet thresholding estimator of Theorem 2.4.4 allows for a quite general dependence structure for the data generating process in theory, where the nonparametric regression problem is asymptotically equivalent to estimating the means of a sequence of independent Gaussian random variables, i.e. the expansion coefficients of γ with respect to an orthonormal basis. It covers, for example, covariance stationary time series $\{\epsilon_t\}$ with short range dependence, i.e. one whose autocorrelation function $\rho(h)$ is absolutely summable with $\sum_{-\infty}^{+\infty} |\rho(h)| < \infty$, and Gaussian innovations. In such cases we would have the sequence of random functions

$$t \mapsto \frac{1}{n} \sum_{i=1}^{\lceil nt \rceil} u_t$$

equivalent to a sequence of Brownian motions $\frac{\tau}{\sqrt{n}} dB_t$ where $\tau^2 = \sum_{-\infty}^{+\infty} |\rho(h)|$ (see [43]).

In simulations, however, we found that estimates of γ in the presence of short range dependence to be noisier than the white noise case. In finite sample, serial correlation may dissipate slowly relative to sample size. As a finite sample adjustment, we therefore consider the situation where the nonparametric regression problem features long range dependent errors. The resulting *level-dependent* wavelet thresholding estimator outperforms the universal wavelet thresholding estimator of Theorem 2.4.4 in simulations, shown in Figure 2.3.

Assumption 4. *In the nonparametric regression problem of Equation 2.5,* u_t *is Gaussian and stationary with long range dependence, i.e. its autocorrelation function* $\rho(h)$ *has sub-hyperbolic decay:* $\rho(h) \approx \frac{A}{|h|^{\alpha}}$ *for some* $0 < \alpha < 1$ *and* $A > 0$.

The trade-off with allowing for long range dependence is that one may need Gaussianity as a distributional assumption.[16] Unlike the short range dependent case, the whitening effect of wavelets, not shared by other basis, is now required to establish asymptotic model equivalence. In this setting, the limit process is now the fractional Brownian motion dB_t^H with Hurst index H, which is determined by $H = 1 - \frac{\alpha}{2} \in (\frac{1}{2}, 1)$. dB_t^H is a mean-zero Gaussian process that behaves like $(\Delta t)^H$ for small time increments Δt, and self-similar in the sense that $B^H(ct)$ and $c^H B^H(t)$ have the same distribution. For $H = \frac{1}{2}$, dB_t^H is the Brownian motion. Unlike the Brownian motion, dB_t^H does not have independent increments in general nor is it a semi-martingale. (The Appendix contains a precise definition of dB_t^H.)

By a Functional Central Limit Theorem for long range dependent processes of [106], the observation and error partial sum processes satisfy

$$n^{1-H}(F_t^{(n)} - \int_0^t \gamma(s)ds) \to \tau B_t^H$$

where the asymptotic variance $\tau^2 = \frac{2A}{(1-\alpha)(2-\alpha)}$. This gives the long-range filtering model

$$dF_t = \gamma(t)dt + \frac{\tau}{(\sqrt{n})^\alpha} dB_t^H.$$

Unlike the case of Brownian motion, where the Iˆto isometry yields i.i.d. Gaussian sequence for any basis of $L^2[0,1]$, the long memory situation require unique properties of wavelets to decorrelate the fractional Brownian motion. By performing a principal component analysis on the reproducing kernel of dB_t^H using wavelets, we obtain a Lévy-Ciesielski-Iˆto type representation (details provided in Appendix):

$$dB_t^H = \sum_{jk} w_{jk} v_{jk}(t)dt,$$

where v_{jk} is approximately orthogonal to $\{\psi_{jk}\}$. The stochastic integrals

$$\int \psi_{jk} dF_t = \int \psi_{jk}\gamma(t)dt + \frac{\tau}{(\sqrt{n})^\alpha} \psi_{jk} dB_t^H$$

give a Gaussian sequence model whose error terms are approximately white noise, with level-dependent error variance.

Theorem 2.4.5. *Denote the empirical wavelet coefficient by* $X_{jk} = \int \psi_{jk} dF_t$. *Then*

(i)

$$X_{jk} = \theta_{jk} + \frac{\tau}{(\sqrt{n})^\alpha} \sigma_j e_{jk},$$

where, at each level j, $\sigma_j = 2^{-j(1-\alpha)}$.

[16] In fact FCLT's for time series with long range dependence, which is all that we require in this discussion, have been established in much more general cases where Gaussianity does not hold. See, for example, [84].

(ii) The random variables $e_{jk} = \frac{1}{\sigma_j}\int_0^1 \psi_{jk} dB_t^H$ *have mean zero, variance* 1*, and are approximately uncorrelated in the sense that* $0 < c_0 \leq Var(e_{jk}|e_{j'k'},\ (j',k') \neq (j,k)) \leq 1$.

It was shown in [45] that a Gaussian sequence model whose noise satisfies the approximate uncorrelated condition of Theorem 2.4.5(ii) has the same asymptotic minimax risk as a model with independent noise. In other words, asymptotically the model is equivalent to the sequence

$$X_{jk} = \theta_{jk} + \frac{\tau}{(\sqrt{n})^\alpha}\sigma_j e'_{jk}$$

where e'_{jk}'s are i.i.d. standard normal. Since for a fixed resolution level j, the equivalent model is a Gaussian sequence model with homoskedastic independent noise for which the universal wavelet threshold estimator achieves minimax risk, as a corollary of Theorem 2.4.4 we arrive at a level-dependent thresholding estimator in the long memory case.

Theorem 2.4.6. *Suppose Assumptions 1(vi), 1(vii) and 4 hold. Let* ψ*, the mother wavelet that generates the wavelet basis used, have* r *vanishing moments and* r *continuous derivatives, where* $r > \max\{1, \eta\}$*. In the nonparametric regression problem of Equation 2.5, let* $R_n(p,q,\eta,L)$ *denote the minimax risk over the Besov ball* $B^\eta_{p,q}(L)$ *and* $\hat{\gamma}_n$ *be the estimate obtained by applying the universal wavelet threshold estimator to each level* j*. Then*

(i)

$$\frac{\sup_{\gamma \in B^\eta_{p,q}(L)} \mathbb{E}[\|\hat{\gamma}_n - \gamma\|^2]}{R_n(p,q,\eta,L)} = O(2\log n + 1),$$

for all $1 \leq p,q \leq \infty$*,* $0 < L < \infty$*, and* $\eta_0 < \eta < r$ *where*

$$\eta_0 = \max\{\frac{1}{p}, 2(\frac{1}{p} - \frac{1}{2})_+\}.$$

(ii) There exists a constant C *such that*

$$Prob\{\|\hat{\gamma}_n\|_{B^\eta_{p,q}} \leq C\|\gamma\|_{B^\eta_{p,q}}\} \to 1.$$

Also by asymptotic equivalence, the (non-adaptive) minimax rate of convergence of the wavelet threshold estimator is therefore the same as that obtained in [46] for the Gaussian sequence model with level-dependent noise

$$X_{jk} = \theta_{jk} + \frac{\tau}{(\sqrt{n})^\alpha}\sigma_j e'_{jk},$$

under the assumption that the time-varying parameter γ is sufficiently regular relative to the correlation structure of the limit process dB_t^H.

Theorem 2.4.7. *(A la Donoho and Johnstone 1998) Under Assumptions 1(vi), 1(vii) and 4, suppose* $\eta + \frac{1}{2} - \frac{1}{p} > \frac{\alpha(2-p)}{2p}$*, then the wavelet level-dependent threshold estimator for the nonparametric regression problem is rate-optimal and its asymptotic risk satisfies*

$$\liminf_{n\to\infty} \sup_{\gamma \in B^{\eta}_{p,q}(L)} \mathbb{E}[\|\hat{\gamma}_n - \gamma\|^2] = O(n^{-r})$$

where $r = 2 \cdot \frac{(\eta + \frac{1}{2} - \frac{1}{p}) \cdot \frac{\alpha}{2}}{\eta + \frac{1}{2} - \frac{1}{p} + \frac{\alpha}{2}}$.

2.4.2. Estimation Error and Asymptotic Rate

This section derives estimating error from our two-step procedure and the resulting asymptotic risk for the time-varying parameter. As one would expect, the rate of convergence is the slower of the two resulting from each step. To have a concrete discussion of asymptotic risk, we consider time-varying parameters which are piecewise Hölder-continuous.

Assumption 5.

(i) There exists a finite partition $0 = t^{(0)} < t^{(1)} < \cdots < t^{(m)} = 1$ *of* $[0,1]$ *such that* γ *satisfies* $|\gamma(t) - \gamma(s)| \leq L|t-s|^{\rho}$ *on each interval* $(t^{(j)}, t^{j+1})$ *for all* $j = 1, \cdots, m-1$*, for some* $\rho > 0$ *and* $0 < L < \infty$*.*

(ii) The error term $\{\epsilon_t\}$ *is conditional heteroskedastic of the form* $\epsilon_t = \sigma(X_t, t)w_t$*, where* $|\sigma(\cdot, t) - \sigma(\cdot, s)| \leq M(\cdot)|t-s|$ *on* $\mathbb{R}^{m+1} \times [0,1]$ *with* $\|M(\cdot)\|_{\infty} < \infty$*.*

(iii) (X_t, w_t) *is* α*-mixing of size* $-\frac{2r}{r-1}$ *for some* $r > 1$*.*

(iv) The kernel K *is Lipschitz continuous.*

From the modelling perspective, there is little loss of generality in specializing Assumption 1(i) to Assumption 5(i).[17] The points of discontinuity $0 = t^{(0)} < t^{(1)} < \cdots < t^{(m)} = 1$ depends on γ and are unknown to the econometrician. Assumptions 5(ii) and (iii) make characterizations of asymptotic distributions possible. We use Assumption 5 in refining consistency of the first step, described in Theorem 2.2.1, to obtain a statement regarding rates of convergence of the asymptotic mean square error in estimating the time-invariant part of the model.[18]

Theorem 2.4.8. *Under Assumptions 1 and 5, the mean square error of* $\hat{\beta}$ *satisfies*

[17] Assumption 5(i) still certainly does not impose any smoothness assumptions on γ. For example, the non-differentiable sample paths of an independent sum of Brownian motion and a finite activity jump process, which are piecewise Hölder with Hölder exponent $\rho < \frac{1}{2}$.

[18] Twice-differentiability of γ would give the standard rate $O_p(n^{-\frac{4}{5}})$ with optimal bandwidth selection (see, e.g. [21]).

$$\|\hat{\beta} - \beta\|_2^2 = O(n^{-\frac{2\rho}{1+2\rho}}),$$

with optimal selection of bandwidth h_n*.*

Since m is finite, the norms $\|\beta\|_2$ and $\|\beta\|_\infty = \max_{1\le j\le m} |\beta_j|$ on $\mathbb{R}^m$ are equivalent. Therefore $\|\hat{\beta} - \beta\|_\infty$ is also of order $O_p(n^{-\frac{\rho}{1+2\rho}})$. So as a corollary of Theorem 2.4.8, the estimation error term in Equation 2.4 satisfies

$$|\frac{\sum_i (\beta_i - \hat{\beta}_i) x_{i,t}}{x_{m+1,t}}| = O_p(n^{-\frac{\rho}{1+2\rho}})$$

for each t. Therefore

$$\frac{Y_t - \sum_i \hat{\beta}_i x_{i,t}}{x_{m+1,t}} = \gamma(t) + \frac{\epsilon_t}{x_{m+1,t}} + O_p(n^{-\frac{\rho}{1+2\rho}}). \tag{2.7}$$

With $\hat{\gamma}_n$ denoting the noisy γ obtained by plugging in $\hat{\beta}$, let $\hat{\hat{\gamma}}_n$ be the wavelet threshold estimator applied to Equation 2.7. By continuity of the wavelet threshold operator,

$$\hat{\hat{\gamma}}_n = \hat{\gamma}_n + O_p(n^{-\frac{\rho}{1+2\rho}}),$$

in $L^2[0,1]$, which gives the following:

Theorem 2.4.9. *Suppose Assumptions 1, 2, 3, and 5 hold. If* $\eta + \frac{1}{2} - \frac{1}{p} > \frac{\alpha(2-p)}{2p}$*, then the wavelet threshold estimator* $\hat{\hat{\gamma}}$ *for time-varying parameter* γ *in the model given by Equation 2.1 has asymptotic risk satisfying*

$$\liminf_{n\to\infty} \sup_{\gamma \in B^\eta_{p,q}(L)} \mathbb{E}[\|\hat{\hat{\gamma}}_n - \gamma\|_{L^2}^2] = O(\max\{n^{-r}, n^{-\frac{2\rho}{1+2\rho}}\})$$

where $r = 2 \cdot \frac{(\eta + \frac{1}{2} - \frac{1}{p}) \cdot \frac{\alpha}{2}}{\eta + \frac{1}{2} - \frac{1}{p} + \frac{\alpha}{2}}$.

In other words, combining the two steps together, one achieves the worst of the two rates from the time-invariant and time varying parts of the model. Similarly, since $\|\hat{\hat{\gamma}}_n\|_{B^\eta_{p,q}} = \|\hat{\gamma}_n\|_{B^\eta_{p,q}} + O_p(n^{-\frac{\rho}{1+2\rho}})$, the no-spurious-jumps result of Theorem 2.4.6(ii) extends to the full model:

Theorem 2.4.10. *Under Assumptions 1, 2, 3, and 5 and in the same notation as Theorem 2.4.6(ii),*

$$Prob\{\|\hat{\hat{\gamma}}_n\|_{B^\eta_{p,q}} \le const \cdot (\|\gamma\|_{B^\eta_{p,q}} + n^{-\frac{\rho}{1+2\rho}})\} \to 1.$$

Estimation Procedure To summarize, our proposed estimator consists of the following two steps:

1 Estimate the time-invariant part of the model consistently by local smoothing kernel methods at a randomly chosen $t_0 \in (0,1)$, as specified by Equation 2.2 and plug in resulting estimates to obtain $\hat{\gamma}$.

2 Apply a non-linear wavelet orthogonal series estimator to obtain $\hat{\hat{\gamma}}$ as described in Section 2.4.1.

2.5. Monte Carlo Simulations

2.5.1. White noise $\{\epsilon_t\}$

We simulate the following model

$$Y_t = \alpha + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \beta_3 x_{3,t} + \gamma(t) x_{4,t} + \epsilon_t, \tag{2.8}$$

where

- The value of the time-invariant parameters are $\alpha = 7, \beta_1 = 5, \beta_2 = -6, \beta_3 = 2$.
- The functional parameter $\gamma : [0,1] \to \mathbb{R}$ is defined by

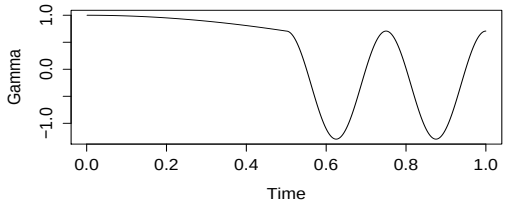
$$\gamma(t) = \begin{cases} \cos \frac{\pi}{2} t & \text{if } t \in [0, \frac{1}{2}] \\ \cos 8\pi t + (\cos \frac{\pi}{4} - \cos 4\pi) & \text{if } t \in [\frac{1}{2}, 1] \end{cases}.$$

  γ is a continuous but not C^1, being not differentiable at $t = \frac{1}{2}$. Figure 2.1a contains a plot of γ. This simulated data generating process has a structural change from low frequency to high frequency oscillation at $t = \frac{1}{2}$.

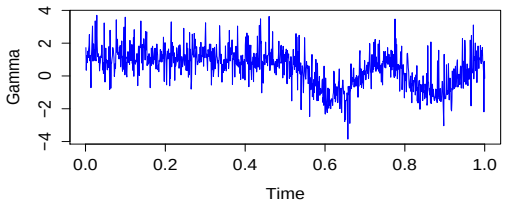
- The regressors and error term have the following independent distributions:
  - $x_{1,t} \sim$ i.i.d. χ^2_2.
  - $x_{2,t} \sim$ i.i.d. $\mathcal{N}(0,1)$.
  - $x_{3,t}$ is a realization of a ARMA$(1,2)$ time series with AR parameters 0.1 and MA parameters $(1,-1)$. The underlying innovation is standard normal white noise.
  - $x_{4,t}$ has i.i.d. distributions which is a χ^2_2 distribution shifted to the right by 1, making it bounded away from zero.
  - $\epsilon_t \sim$ i.i.d. $\mathcal{N}(0,4)$.

The sample size used is $1,000$. We now describe a typical estimate. Using the Epanechnikov kernel, the estimates of time-invariant parameters are

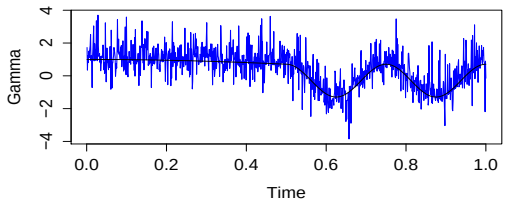
$$\hat{\alpha} = 6.640, \hat{\beta}_1 = 5.058, \hat{\beta}_2 = -6.082, \hat{\beta}_3 = 2.023.$$



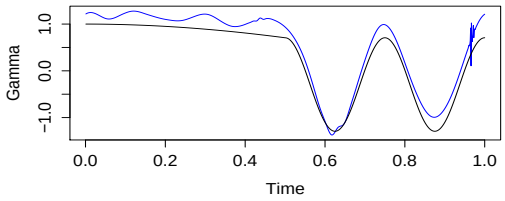
(a) The functional parameter γ.



(b) Z_t obtained by plugging in consistent estimates of time-invariant parameters.



(c) Z_t in comparison with true γ.



(d) Estimate $\hat{\gamma}$ given by the universal wavelet threshold estimator.

Figure 2.1: Estimate of γ by universal wavelet thresholding from a typical realization of data generated by the model of Equation 2.8. The errors $\{\epsilon_t\}$ in the simulated DGP is i.i.d. $\mathcal{N}(0, 4)$ white noise.

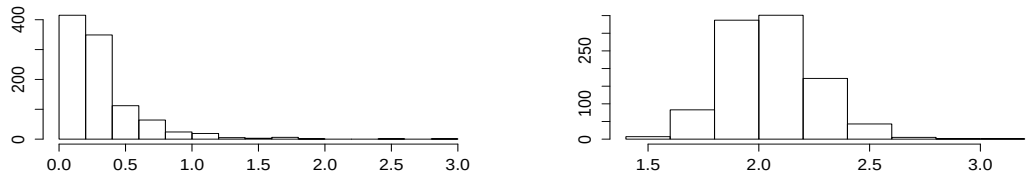
Plugging in gives the noisy version of γ to be wavelet-thresholded, shown in Figure 2.1c,

$$Z_t = \frac{Y_t - \hat{\alpha} - \hat{\beta}_1 x_{1,t} - \hat{\beta}_2 x_{2,t} - \hat{\beta}_3 x_{3,t}}{x_{4,t}}.$$

Even though Z_t is very noisy compare to γ, wavelet thresholding removes most of the noise, with estimate shown in Figure 2.1d. The average mean square error (AMSE), the empirical counterpart to mean integrated square error (MISE), in this case is

$$\frac{1}{n}\sum_t (\hat{\hat{\gamma}}(t) - \gamma(t))^2 = 0.02.$$

On the other hand, the L^2-norm of γ is $\frac{1}{n}\sum_t \gamma(t)^2 = 0.7$, making the relative error approximately 4.2%. This is in spite of a very large noise-to-signal ratio of 2.041884, as measure by the ratio of empirical variances of (Z_t) over that of $\gamma(t)$. Relative errors from a Monte Carlo simulations of $1,000$ repetitions are shown in Figure 2.2, with mean of approximately 6% and standard deviation 0.05. From the same $1,000$ simulations, the noise-to-signal variance ratios have a relatively large mean of 2.00 and standard deviation 0.096.



(a) Relative errors measured by MISE from 1000 simulations.

(b) Noise-to-signal ratio from 1000 simulations.

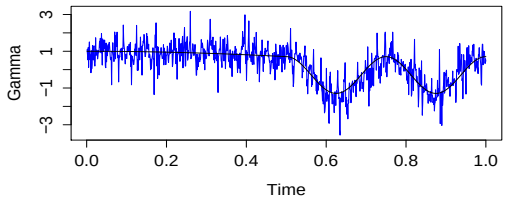
Figure 2.2: Monte Carlo simulation of 1000 repetitions. The average relative error is approximately 6%, which is small considering the amount of noise faced by the estimator.

2.5.2. Short-range dependent $\{\epsilon_t\}$

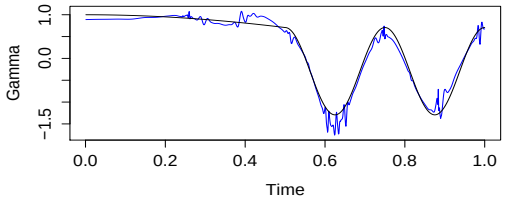
As described in Section 2.4.1, in theory asymptotic minimaxity of the universal threshold estimator in the nonparametric regression problem extends to models with short range dependent error $\{\epsilon_t\}$ of Equation 2.8. However, in simulations we found that the universal threshold estimator gives somewhat noisy estimates of γ when $\{\epsilon_t\}$ has short range dependence. Figure 2.3b shows one such estimate. The simulated DGP satisfies all previous conditions except $\{\epsilon_t\}$ is now an AR(1) time series with AR parameter 0.5. Noticeably more noise survives universal thresholding than in the white noise case. We view this as empirical evidence that while the observation partial sums processes still converge to the Brownian motion, serial dependence slows down the speed of convergence considerably. In comparison, applying the level-dependent threshold estimator specified in Theorem 2.4.6 to the same set of data gives an visually improved estimate, which is shown in Figure 2.3c. Figure 2.4 shows a results from $1,000$ simulations using the level dependent threshold estimator; Figure 2.4a shows the distribution of relative errors and Figure 2.4b shows the distribution of the noise-to-signal ratio faced by the wavelet estimator in the de-noising problem.

2.5.3. Small sample

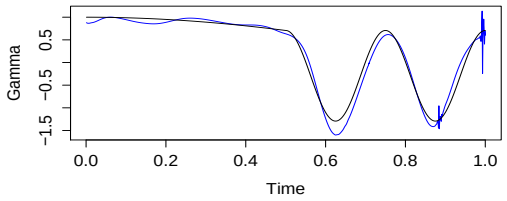
Figure 2.5 contains simulation result of $1,000$ repetitions with sample size reduced from $1,000$ to 250. For comparison, Figure 2.5b shows the level-dependent threshold estimate and Figure 2.5c shows the universal threshold estimate for the same realization of Figure 2.5a . The level-dependent threshold estimator is used, as an empirical adjustment. As expected, the quality of estimates worsens with smaller sample size but it is still acceptable in our view.



(a) Noisy γ in comparison with true γ.



(b) Universal threshold estimate of γ when $\{\epsilon_t\}$ has short-range dependence.



(c) Level-dependent threshold estimate of γ.

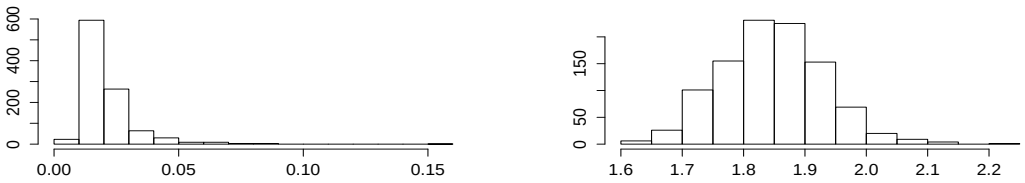
Figure 2.3: The wavelet universal and level-dependent threshold estimators applied to data with short-range dependent errors.

2.5.4. Discontinuous γ

Consider now a γ that undergoes simultaneously smooth structural change and abrupt structural break at $t = \frac{1}{2}$, plotted in Figur 2.6a:

$$\gamma(t) = \begin{cases} \cos \frac{\pi}{2} t & \text{if } t \in [0, \frac{1}{2}) \\ \cos 8\pi t + 3 & \text{if } t \in [\frac{1}{2}, 1] \end{cases}.$$

With a sample size of $1,000$ and short range dependent errors, typical estimates using the level-dependent threshold estimator and the universal threshold estimator are shown in Figures 2.6c and 2.6d, respectively. Both capture well the true jump at $t = \frac{1}{2}$, although the universal threshold estimate is visually more noisy as expected. It is the level-dependent estimator that is used in the $1,000$-repetition Monte Carlo simulation whose results are shown in Figure 2.7. The average relative error of approximately 2.13% is approximately the same as the 2.08% obtained for a γ that is continuous, in Figure 2.4. The noise-to-signal ratio is approximately the same in the two cases. Thus the performance of the wavelet estimator is unaffected by possible discontinuities in γ. Whereas other estimators may breakdown in the presence of discontinuity, the wavelet estimator can actually exploit the additional spatial inhomogeneity in γ. With a smaller sample size of 250, the wavelet estimate still retains the essential features of true γ as shown by results in Figure 2.8. Again comparing to the case where γ is continuous with the same sample size of 250, discontinuity



(a) Relative errors, mean = 0.02083235 × 100%, sd = = 0.01107202.

(b) Noise-to-signal ratio, mean = 1.846658, sd = 0.08510039.

Figure 2.4: Monte-Carlo simulation of 1,000 repetitions with short range dependent errors.

in γ causes no discernible degradation on the quality of the estimate, with the noise level being comparable.

2.5.5. Comparison with other methods

Local smoothing While local kernel regression theory does not accommodate discontinuities, in an empirical setting it may nevertheless be applied in *ad hoc* manner to the model of Equation2.1.[19] This is done for the model of Equation 2.8 with Monte Carlo results shown in Figure 2.9. As measured by MISE, the results are comparable with the wavelet estimator, with wavelet estimator slightly better. Figures 2.11a and 2.11b show the two estimates for one sample.[20]

More importantly, when applied to a model that violate smoothness assumptions, the local smoothing estimator may result in false discovery of time trend in a parameter when none exists. The time-varying parameter γ in the simulated model of Equation 2.8 has an abrupt jump at $t = \frac{1}{2}$ with the other parameters being constant. The range of vari-

[19] In principle, one may measure the performance of a local smoothing estimator by MISE rather than pointwise convergence. We do not know of any such theoretical consideration in the literature.

[20] We used the Epanechnikov kernel. Bandwidth h is selected between 0.02 and 0.2 in steps of 0.02 using the Akaike Information Criterion

$$AIC(h) = \log(RSS) + \frac{n + tr(S^*)}{n - (tr(S^*) + 2)}$$

where $RSS = \sum_{t=1}^{n} Y_t - \hat{Y}(h)_t$ and S^* is the linear smoothing operator in the kernel literature (see, for example, [70]).

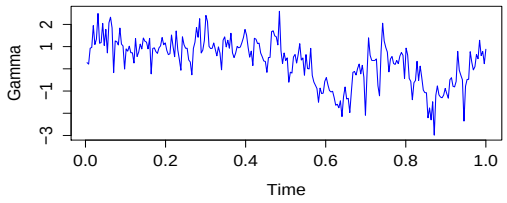
In our Nadaraya-Watson case, $S^*(h) = [S_1(h), \cdots, S_n(h)]'$, where

$$S_t(h) = A_t(h)' X_t,$$

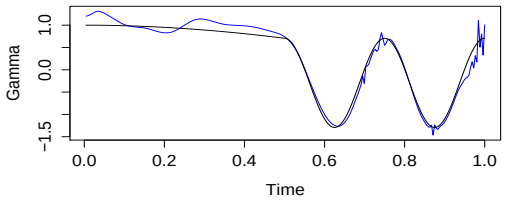
and

$$A_t = (X' W_t(h) X)^{-1} X' W_t(h), \quad W_t = diag(K_h(t_1 - t), \cdots, K_h(t_n - t)).$$

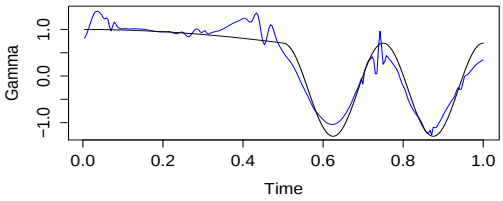
In producing Figure 2.9a, this bandwidth selection procedure is applied for each of the 1,000 realizations.



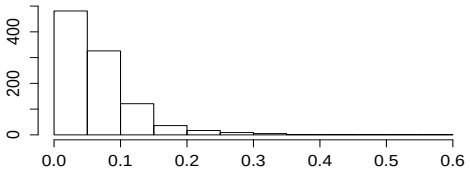
(a) One realization of noisy γ with AR(1) $\{\epsilon_t\}$.



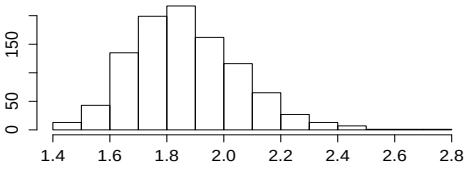
(b) Corresponding level-dependent threshold estimate of γ with AR(1) $\{\epsilon_t\}$.



(c) Corresponding universal threshold estimate of γ with AR(1) $\{\epsilon_t\}$.



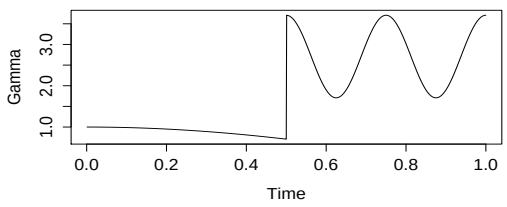
(d) Relative errors, mean $= 0.07042005 \times 100\%$, sd $= 0.057314$.



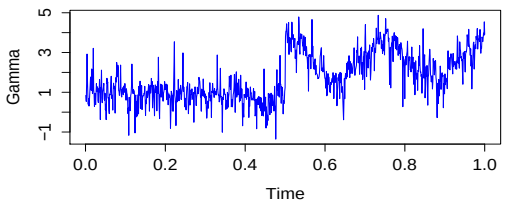
(e) Noise-to-signal ratio, mean $= 1.868884$, sd $= 0.1902459$.

Figure 2.5: Monte-Carlo simulation of 1000 repetitions with short range dependent errors. Sample size reduced from $1,000$ to 250.

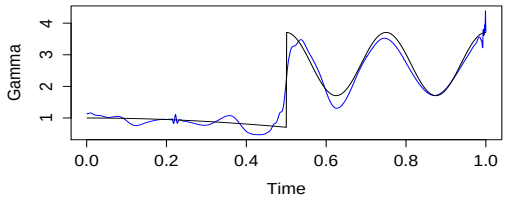
ability of γ, $|\max_{t\in[0,1]}\gamma(t) - \min_{t\in[0,1]}\gamma(t)|$, coincides with the magnitude of the jump $|\lim_{t\to\frac{1}{2}^+}\gamma(t) - \lim_{t\to\frac{1}{2}^-}\gamma(t)| = 3$. However, the kernel estimates of the time-invariant parameters can exhibit a range of variability similar to, or possibly exceeding, the estimated γ or true γ. Figures 2.11c, 2.11d, 2.11e, and 2.11f show two estimates for α and β_2 each. Figure 2.11c shows that, as one might expect, a distortion of estimate around $t = 0.5$ where the data generating process is discontinuous. In addition, even when away from the point of discontinuity, local smoothing may suffer distortion, as shown more clearly around $t = 0.8$ in Figure 2.11d. When compared with estimates of γ in Figure 2.11e and 2.11f, the erratic estimates would suggest the misleading conclusion that α and β_2 vary in time as much as γ. Monte Carlo results for the variability of time-invariant parameter estimates are shown in Table . The reason lies in the fact that any optimal bandwidth selection (e.g. AIC, BIC, or



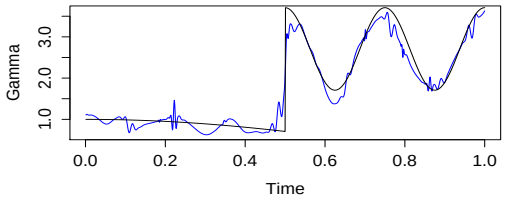
(a) γ with discontinuity.



(b) One realization of noisy γ with AR(1) $\{\epsilon_t\}$.



(c) Level-dependent threshold estimate of γ with AR(1) $\{\epsilon_t\}$.



(d) Universal threshold estimate of γ.

Figure 2.6: Discontinuous γ with short range dependent errors.

cross validation) is necessarily a global procedure. Discontinuity results in misspecification that distorts such attempts to achieve optimal bias-variance trade-off. Along with absence of theoretical justification, this shows that the added generality of possibly multiple time-varying parameters of the local smoothing approach does not extend a a model with abrupt parameter changes.

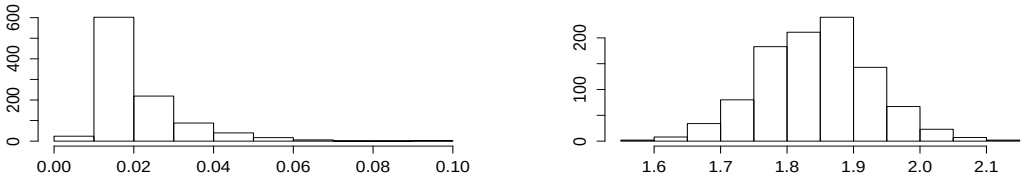
Splines Another commonly considered nonparametric techniques is the use of splines. Similar to local smoothing, using splines presupposes that the underlying parameter is smooth.[21] As in the kernel case, we compare an *ad hoc* application of splines with the wavelet estimator.

We follow standard practice in spline literature by using cubic splines ϕ_k, placing a knot at each data point, and applying a roughness penalty.[22] Empirically, expand γ into $\gamma(t) \approx \sum_{k=1}^{K} \phi_k(t)$. The working model is then

$$Y_t \approx \alpha + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \beta_3 x_{3,t} + \sum_{k=1}^{K} \phi_k(t) x_{4,t} + \epsilon_t,$$

The estimates and the length of expansion K are determined by penalized least squares criterion

[21] Sobolev spaces are, for example, function spaces commonly associated with splines.
[22] See, for example, [99].



(a) Relative error, mean = 0.02125429 × 100%, sd = = 0.01088821.

(b) Noise-to-signal ratio, mean = 1.84448, sd = 0.08350869.

Figure 2.7: Monte-Carlo simulation of 1,000 repetitions with short range dependent errors and γ with discontinuity.

$$\sum_{t=1}^{n}(Y_t - \mathbf{b}'\tilde{X}_t)^2 + \lambda \mathbf{b}' R \lambda \mathbf{b},$$

where $\tilde{X}_t = (1, x_{1,t}, x_{2,t}, x_{3,t}, \phi_1(t), \cdots, \phi_K(t))'$, and R is the roughness penalty matrix of order K defined by $R = diag(I_4, [\int \phi_i'' \phi_j'']_{1 \leq, i,j \leq k})$. The roughness penalty coefficient λ is in turn selected by the generalized cross validation criterion

$$GCV(\lambda) = \frac{n \cdot RSS}{(n - df(\lambda))^2}$$

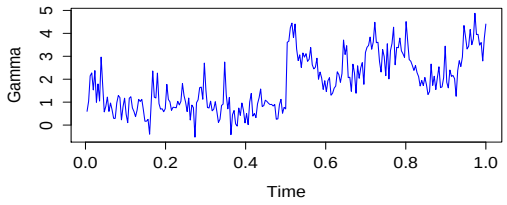
where the effective sample size $df(\lambda) = trace\tilde{X}(\tilde{X}'\tilde{X} + \lambda R)^{-1}\tilde{X}'$, from 0.005 to 0.01 in steps of 0.0005.

Figures 2.12b and 2.12c show two typical spline estimates, with visible severe over-smoothing. Smooth splines are unable to capture discontinuity in the parameter. Figures 2.12d shows the results of Monte Carlo with 1,000 repetitions, which are clear inferior to that of wavelet estimator.[23]

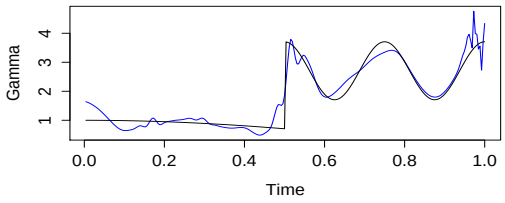
2.6. Time-Varying Systematic Risk

As an empirical exercise, we apply our methodology to estimate time-varying systematic risk of securities. Our model specializes to a capital asset pricing model (CAPM) with time-varying beta. Ample empirical evidence exists for the inadequacy of the static CAPM model and the asset pricing literature contains extensive investigations of the CAPM model with time-varying betas. [35] extracts the component of the unsystematic risk due to nonstationary beta under the random walk assumption on beta. One can also adopt a model

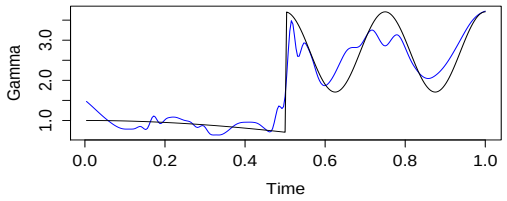
[23]Splines are also computationally much more expensive. In our case, the 1,000 repetitions used in producing Figure 2.12d took approximately 90 hours. Similarly, the 1,000 repetitions for Figure 2.9b using local smoothing required approximately 47 hours. In contrast, the same simulation of the wavelet estimator can be done in well under 1 hour.



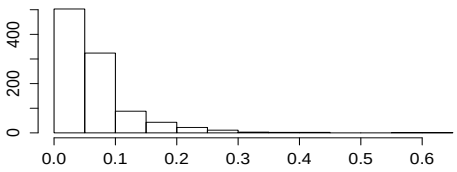
(a) One realization of noisy γ with AR(1) $\{\epsilon_t\}$.



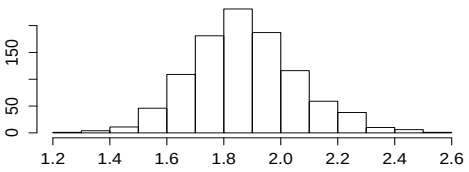
(b) Corresponding level-dependent threshold estimate of γ with AR(1) $\{\epsilon_t\}$.



(c) Corresponding universal threshold estimate of γ with AR(1) $\{\epsilon_t\}$.



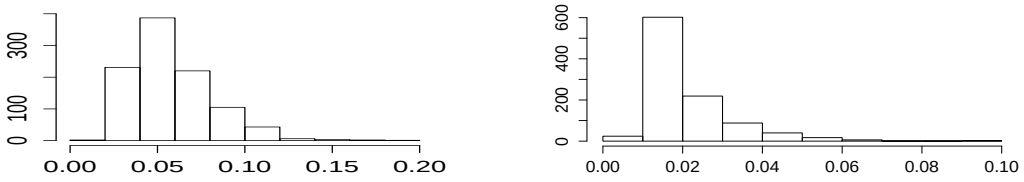
(d) Relative errors, mean $= 0.07014388 \times 100\%$, sd $= 0.05950948$.



(e) Noise-to-signal ratio, mean $= 1.87364$, sd $= 0.1876208$.

Figure 2.8: Monte-Carlo simulation of $1,000$ repetitions with short range dependent errors with discontinuous γ. Sample size reduced from $1,000$ to 250.

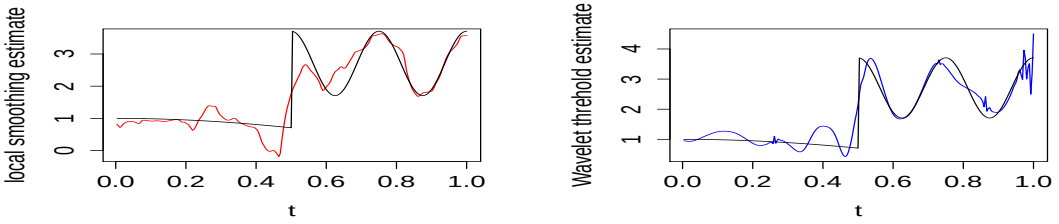
where the returns follow a time series specification as in [24]. Time series specifications carry the implication that changes in the joint distribution of all the securities in the market are captured by the parameters within the model. This may prove to be inadequate, either due to simple misspecification or that the time series parameters themselves may vary in time. [59] shows that the latter case does indeed arise. It was shown that when both the static model and a parametric time-varying beta model are misspecified, there are situations where the static model outperforms the parametric time-varying beta model in pricing risk. Non-constancy of time-series parameters may result in sufficiently severe misspecification to overcome any additional flexibility gained by allowing for parametric time-variation. Modelling beta as a arbitrary, not necessarily smooth, function of time removes such misspecification issues. Also of interest is a model allowing for time-varying



(a) Local Smoothing. Relative errors, mean = $0.0576916 \times 100\%$, sd = 0.0231855.

(b) Wavelet Estimate. Relative error, mean = $0.02125429 \times 100\%$, sd = 0.01088821.

Figure 2.9: Monte-Carlo comparison of 1000 repetitions with short range dependent errors. Sample size = $1,000$.



(a) Local Smoothing.

(b) Wavelet Estimate.

Figure 2.10: Comparison of Local Smoothing and Wavelet Estimates, Sample size = 250.

alpha rather than beta. For example, [49] shows that the contribution of time-varying beta to variation in returns is small compared to the contributions of changing risk premia.

We apply the wavelet estimator to daily returns of 10 U.S. stocks in the year 2010, from January 1 to December 31, in the time varying beta CAPM specification

$$r_t = \alpha + \beta_t m_t + \epsilon_t, \tag{2.9}$$

where r_t is return of a given security and m_t is market return, proxies by that of S&P 500 index. All stocks are inluded in the S&P 500 index. Sample size is 255 trading days. Results are shown in Figure 2.13. Rather than requiring months or years of past return data, Equation 2.9 is a high frequency formulation that captures daily variation in systematic risk.[24]

[24]In contrast, for example, the Bloomberg Professional service computes beta using weekly data over a two year period using a constant beta regression, while Yahoo! Finance uses monthly data over a three year period. This may be of limited relevance for the increasing number of market participants who have a trading horizon that is much shorter, e.g. intraday.

Table 2.1: Spurious Time-Variability

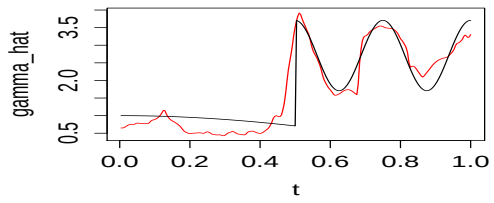
Sample size $n = 250$	$\hat{\alpha}$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
Mean of max - min	4.96	1.34	2.69	1.36
Mean of estimates at $t = 0.5$	6.85	5.02	-5.99	2.00
sd of estimates at $t = 0.5$	2.45	0.63	1.15	0.62
Mean of estimates at $t = 0.8$	7.00	4.98	-5.97	1.99
sd of estimates at $t = 0.8$	1.21	0.29	0.52	0.29
Sample size $n = 1,000$	$\hat{\alpha}$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
Mean of max - min	4.31	1.08	2.35	1.13
Mean of estimates at $t = 0.5$	6.98	5.01	-5.98	1.99
sd of estimates at $t = 0.5$	2.31	0.50	0.96	0.51
Mean of estimates at $t = 0.8$	7.03	4.99	-6.01	1.99
sd of estimates at $t = 0.8$	0.87	0.20	0.39	0.23

Monte Carlo of kernel estimates are carried out in 1,000 repititions for sample sizes of 250 and $1,000$. The rows labeled "Mean of max - min" show the Monte Carlo estimates of estimate variability of time variant part of the model when using kernel smoothing. The time variability of true γ is $|\max_{t\in[0,1]}\gamma(t) - \min_{t\in[0,1]}\gamma(t)| = 3$. For both sample size of 250 and 1,000, all estimates suggest time-varying behavior from constancy when none exists.

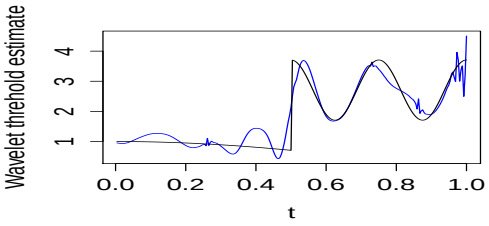
2.7. Summary

In this chapter we provide a framework for functional estimation of a time-varying parameter in a linear regression model. Leaving the paradigm of pointwise estimation, we estimate the parameter over the entire observed period simultaneously. Our approach is in enabled by incorporating the use of wavelets. Wavelets stand out as the only family of basis functions that can both efficiently encode spatial inhomogeneity and whiten serial correlated time series. Both properties are exploited in our methodology. By transforming the estimation problem into one of estimating wavelet coefficients, we make use of the fact that the wavelet decomposition of spatially inhomogeneous functions are concentrated at a few relatively large coefficients. Furthermore, the estimation problems are statistically independent across resolution levels even when the error terms feature serial correlation. The allowance for jumps, abrupt breaks, and smooth trends of any type is in contrast with local smoothing methods, which imposes smoothness assumptions. Local smoothing also may suggest time-varying behavior when none exists. Wavelets also offer significant computational advantages. In this more general setting with respect to time-varying behavior, important characteristics of the parameter such as the locations of possible discontinuities and different types of time trending behavior are captured by the wavelet estimate. Possible

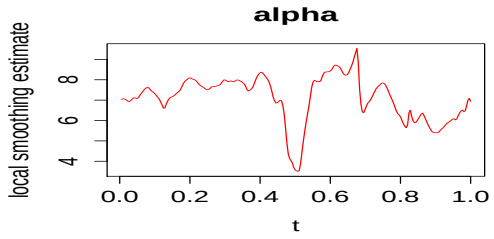
empirical applications besides the one considered in Section 2.6 include using our technique to investigate structural breaks in financial market quality (e.g. price impact, liquidity, and informational efficiency) during an events such as May 6 2010 Flash Crash and the October 15th 2014 bond market flash crash. The latter event lasted approximately 15 minutes during which the yield on the 10-year Treasury plunged an unprecedented amount. Techniques designed for smoothly varying parameters would not be suitable in such a setting where there is clear evidence of abrupt changes.



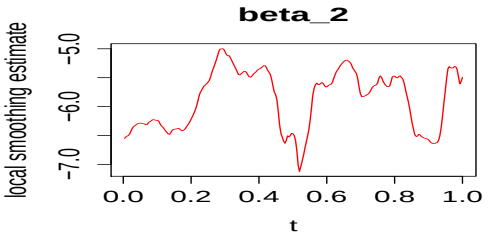
(a) Local Smoothing, optimal bandwidth $h = 0.06$.



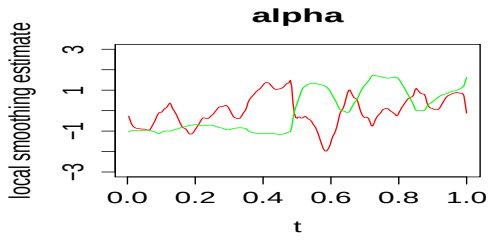
(b) Wavelet Estimate.



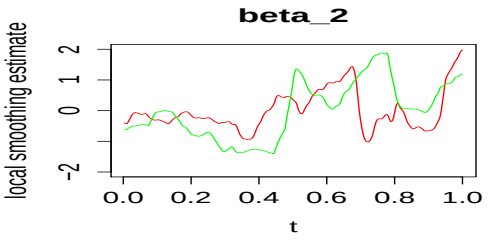
(c) Local Smoothing estimate of α.



(d) Local Smoothing estimate of β_2.



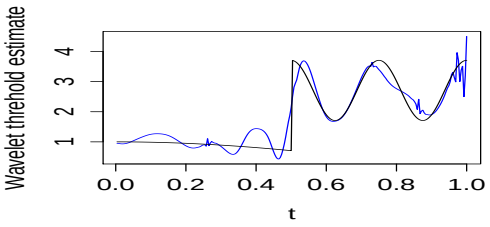
(e) Spurious time trend in $\hat{\alpha}$.



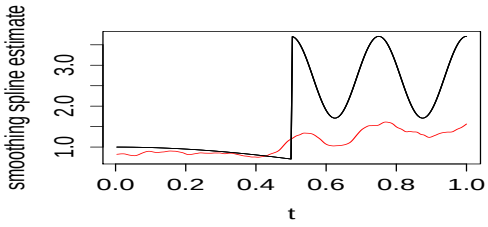
(f) Spurious time varying trend in $\hat{\beta}_2$.

Figure 2.11: Spurious Time Trend from Local Smoothing, Sample size = 250.

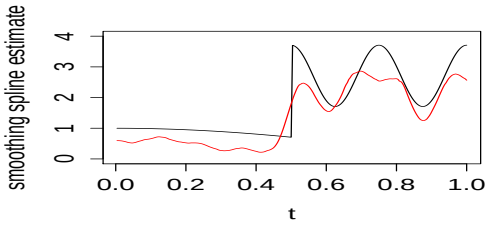
Figures 2.11c and 2.11d show the kernel estimates of a time-invariant parameters α and β_2 respectively, with spurious time-varying behavior. Figures 2.11e and 2.11f compares kernel estimates of the time-invariant parameters (shown in red) with that of the time-varying parameter γ (shown in green).



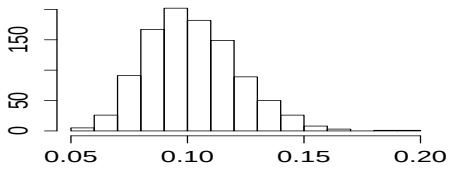
(a) Wavelet Estimate.



(b) Smoothing Spline Estimate, roughness penalty $\lambda = 0.0035$.



(c) Smoothing Spline Estimate, roughness penalty $\lambda = 0.002$.



(d) mean $= 0.1022744 \times 100\%$, sd $= 0.01956175$.

Figure 2.12: Comparison of Spline and Wavelet Estimates, Sample size = 250.

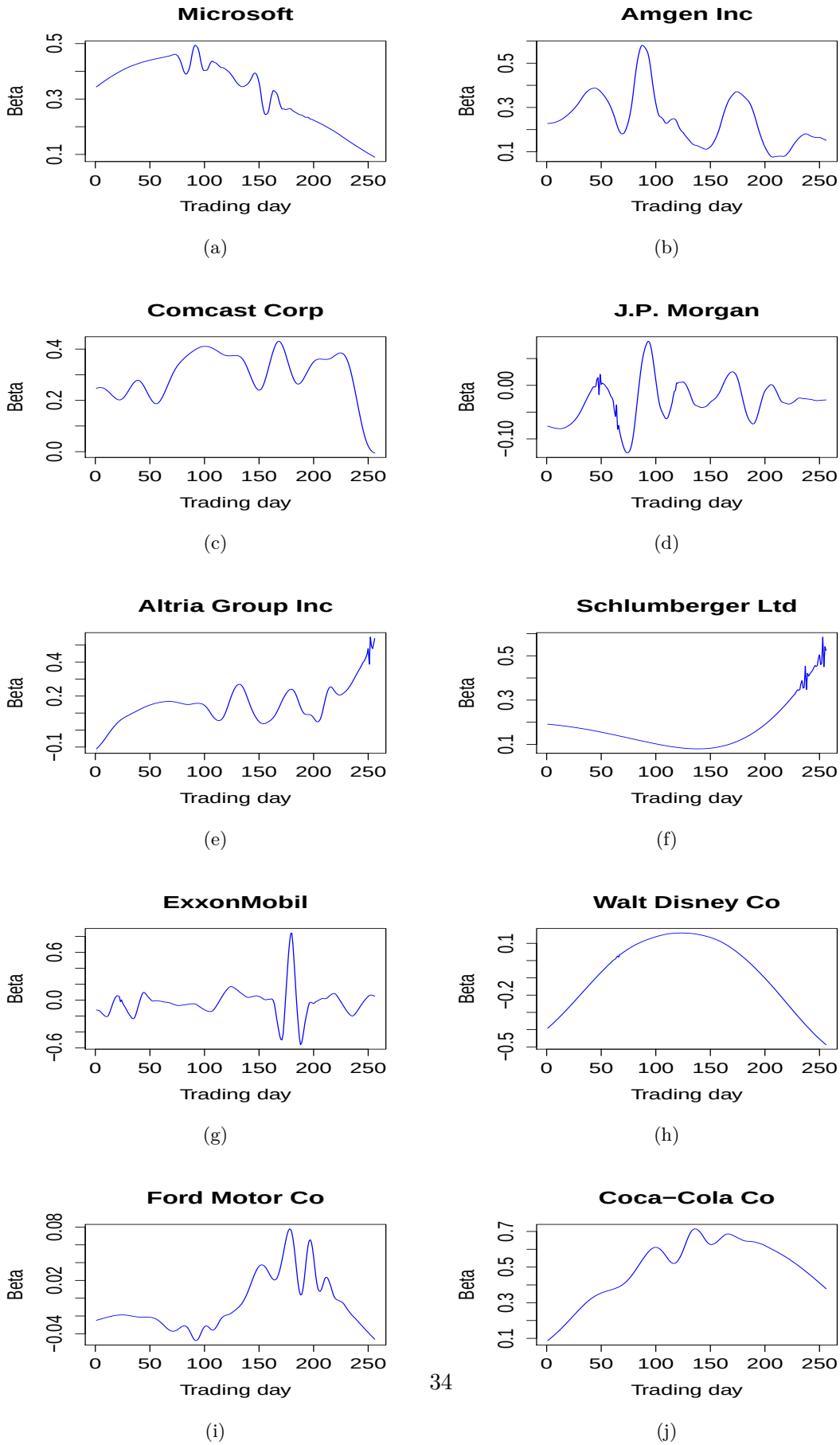


Figure 2.13: Time-varying beta of 10 U.S. stocks, estimated using 2010 daily returns.

Chapter 3

High Frequency Microstructure

3.1. A Need for Re-examination

Market microstructure literature is still coming to grasp with rapid progression of the modern market toward higher speed and lower latency, with particular impetus stemming from events such as the May 6, 2010 flash crash, and the October 15, 2014 bond market flash crash. Much of this fundamental transformation is driven by the specie of traders—more precisely trading algorithms—who operate at the edge of the evolving speed envelope. As pointed out by [92], microstructure research is confronted with basic questions ranging from the nature of information, scope of market data, and suitability of existing empirical measures and techniques in analyzing market quality. The literature lacks agreement regarding the principal issue on whether algorithmic trading has positive impact on market quality. Empirical studies offer opposing conclusions. Previous studies based on econometric approaches present results favorable to algorithmic trading. For example, based on a regression of spread-based measures on a proxy for algorithmic trading activity, [71] conclude that that algorithmic trading enhances the informativeness of quotes and improves liquidity in NYSE limit order books. [26] show similar findings for NASDAQ. In estimating an econometric model of price using NASDAQ data, [69] concludes that algorithmic trading activity contributes more to the efficient martingale component and also lowers short-term volatility and decreases spreads. On the other hand, putting aside econometric evidence, tactics used by algorithmic traders for market manipulation have also been documented. For example, [20] discusses “stuffing", “smoking", and “spoofing" on platforms including NYSE and NASDAQ.[1] In this chapter we offer, in a single setting on one trading platform where these two apparently contradictory findings come together. We reconcile the findings by arguing that the apparent improvement in market quality measures is observed precisely because of certain predatory trading behavior by algorithmic traders.

[1]For example, “spoofing" is a bait-and-switch tactic which involves submitting and canceling orders with no intention of execution, with the goal of swaying prices in a favorable direction.

We make use of a microstructure event that intensified algorithmic trading activity in a main foreign exchange (FX) interdealer trading venue. As an asset class, the FX market is the largest market in the world. The spot FX market daily turn over, 1.5 trillion USD in 2013, is approximately 7 times that of equity markets.[2] A significant portion of this volume, estimated at 35%, are interdealer trades between large financial institutions. Given the immensity of the FX market and its potential impact, how microstructure impacts informational relationship between FX market participants is an important question. On March 7, 2011, Electronic Broking Services (EBS), one of the two main interdealer FX trading platforms, implemented a change of tick size, i.e. minimum price increment, from four to five decimal places —from *pip* to *decimal pip* pricing, in FX market vernacular.[3] Using this microstructure change which increased algorithmic trading activity, we study the changes in measures of adverse selection and analyze trader behavior that using high-frequency limit order book evolution.

As our data is at 100 millisecond frequency, we treat the exchange rate as a continuous-time semimartingale rather than a discrete time series. From the microstructure point of view, order flow represents noisy information flow, from liquidity consumers/market takers to liquidity providers/market makers (see. e.g. the seminal papers [83], [62] and their descendants). In a limit order book market, the jump component of the semimartingale price process directly captures the most conspicuous part of market taker's liquidity consumption. Jumps in prices are caused by large market orders.[4] Serial correlation properties of jump component of price therefore indicate the information content of large order flows. For example, staggered jumps in the same direction arriving in a clustered manner show that the market orders responsible for the jumps are driven by market taker's private information. Conversely, a jump time series which is white noise means large market orders convey no information. Such jumps are caused by market orders placed for exogenous liquidity reasons. Analyzing the price process spanning the entire two years of 2010 to 2011, we find that the jump time series reject strongly the null hypothesis of no serial correlation before tick size change and there is no serial correlation in large order flow after adoption of decimal pip. Similar to [69], we find that price becomes more martingale-like as algorithmic trading activity intensifies.

In addition to the exchange rate process, the econometric component of our analysis considers the entire limit order book during the same period. Decomposing the quoted spread into components due to transaction cost, inventory risk, and adverse selection shows

[2]See the Bank for International Settlements (BIS) report: `http://www.bis.org/publ/rpfx13fx.pdf`. The spot market makes up 37% of the global FX market. Other currency related instruments include FX swaps, forwards, currency swaps, and options.

[3]"Pip" is abbreviation for Price Increment Point.

[4]To the best of our knowledge, price jumps caused by market makers adjusting quotes by large discrete increments –known as price "gapping"– are rare occurrences for electronic order book markets in general, apart from singular catastrophic events like the Flash Crash. We do not observe any price gapping in our data.

that the adverse selection component decreases after tick size change. We also consider an empirical proxy of adverse selection defined using the effective spread and realized spread. We find clear negative level shift in the proxy time series after tick size change. Therefore, adverse selection measures decrease unconditionally for market makers, not just conditional on large orders.

In line with previous studies, uniform improvement in market quality measures might suggest that the speed of algorithmic traders improves liquidity and efficiency of prices, resulting in a positive shift in market quality. On the other hand, we also observe emergence of high-frequency predatory trading behavior by algorithmic traders after tick size change. We show that algorithmic traders exploit trading options made avaiable by tick size change at the expense of other traders. After tick size change, the shape of the limit order book exhibits pronounced price clustering at previous tick size. Our data suggest strongly that this peculiar shape is due to manual market makers' failure, or inability, to adapt to new decimal tick size. The refusal of manual market makers to adapt to new decimal tick size accentuates the easiness with which algorithmic market makers can employ the queue jumping strategy already made available by tick size change.

Based on exogeneity of tick size change in this setting, our results is a case study on how a microstructure event affect the informational relationship between market participants and how algorithmic trader behavior may distort market quality measures. This is the first paper to reconcile the apparent improvement in market quality due to the presence of algorithmic traders and their predatory market making activity.

The rest of this chapter is organized as follows. Section 3.2 reviews related literature covering the three strands of technical part of our analysis: high frequency continuous-time econometrics, limit order book econometrics, and tick size change. Section 3.3 summarizes relevant institutional details of the interdear FX market. Section 3.4 gives a description of EBS order book data. Section 3.5 considers the structural change in the jump component of the price processes before and after tick size change. En route to our analysis of serial correlation of jumps, we confirm stylized facts regarding the foreign exchange market. Section 3.6 discusses components of the quoted spread before and after tick size change. In addition to adverse selection, we confirm the market maker's inventory risk decrease after tick size change as expected. Section 3.7 discusses the realized spread before and after tick size change and its implication on information asymmetry. Section 3.8 explains the observed reduction in adverse selection by presenting an analysis of the shape of the order book and its implications regarding market participants behavior.

3.2. Related Literature

As appropriate for our high-frequency data, we model the price process by a continuous-time semi-martingale.[5] The econometric analysis of such processes is first considered by [15] which separates quadratic variation into its continuous and jump components. The quadratic variation is estimated by realized variance and integrated volatility estimated by realized bipower variation, with the difference between the two provides a consistent estimate of the jump contribution to price variation. [16] provide an asymptotic distribution theory to construct non-parametric tests for the presence of jumps; this is the test we use in testing for jumps. Finite sample refinements have been offered by [78] and [17]. Studies that apply this methodology to data (see e.g. [6], [7], and [8] in the context of equity markets) have been confined to jump and volatility estimation. Rather than merely detecting jumps, we focus on the informational implications of jump behavior. We do, however, confirm stylized facts regarding seasonality of high frequency markets en route.

In dealing with the trade-off between approaching the high-frequency limit and facing contamination by microstructure noise, we adopt the same approach of [6] in comparing the difference between realized volatility and bipower variation, which consistently estimates the jump contribution to price variation in the absence of microstructure noise, across different sampling frequencies. Our sampling frequency is every 30 seconds; we are not aware of any other studies that samples the price process at a frequency higher than every 5 minutes. An alternative approach introduced by [80], [98] and [97] is to exploit the data at the highest frequency but uses local pre-averaging to produce noise reduced observations. [36] construct noise –and outliers– robust estimator using pre-averaging method. Rather than merely seeking to escape the frequency zone occupied by microstructure noise, local pre-averaging requires statistical assumptions on the nature of microstructure noise that might not be empirically justified for the EBS market we consider.

Microstructure theory divides the quoted spread into three components: expected loss due to adverse selection ([62], [41]), inventory risk ([105], [5], [73]) and transaction cost ([102]). The trade indicator model of [76] has been used extensively in analyzing components of the spread, in the context of equity markets. For example, [108] examine the NYSE and NASDAQ spreads for the same stocks. An alternative way to estimate adverse selection in the market is [48] (see also [47]), probability of informed trading model. Their volume-based approach bypasses the spread and is suitable for markets where algorithmic brokerage is prevalent. Considering the make-up of marker participants in the inter-dealer FX market, we use the [76] model in decomposing the spread.

The microstructure perspective on foreign exchange markets originated with [89], which tests validity of structural microstructure models and confirms both the informational and inventory-control aspects of market maker in reaction to incoming order flow. In this paper,

[5] Asset pricing theory postulates that in an arbitrage-free market, prices necessarily follow a semi-martingale.

we focuses on the informational aspect. [94], using the variance decomposition methodology of [68], showed that adverse selection contributed to 60% of quoted spread in one week USD/DEM data from Reuters D2002-2 dealing system. [23] find also that adverse selection component makes up for a large portion of spread in the trade indicator model of [76], although dealer's own prices do not reflect inventory control effort. In this paper, we investigate the change of adverse selection component due to a specific microstructure event, namely tick size change.

Tick size drew the attention of microstructure literature as the minimum price increment in the US equity markets moved from one-eighth to one-sixteenth, and finally decimal pricing (see, for example, [11], [63], and [104]). Extensive studies have been done on the relationship between tick size and informational structure for equity markets, with a diverse array of conclusions reached. [67] advances empirically the position that small tick size benefits professional traders at the expense of large order traders and public traders who use limit orders. [60] examines NYSE data from approximately two months before and after decimalized pricing of January 29, 2001. Also using methodology of [76], they found that adverse selection component of the spread did not change significantly across decimalization, with the possible explanation that pre-decimalization spread was artificially too large, allowing dealers to charge excessive order processing cost. To the best of our knowledge, there is no discussion in the literature on effect of tick size on adverse selection in the FX market. We provide a multi-faceted analysis on effect of tick size change on adverse selection and an empirical explanation, based on a detailed examination of the limit order book, for the observed results.

Our analysis of the limit order book shows that price clustering becomes a distinguishing characteristic of manual trader behavior only after tick size change. The price clustering phenomenon has previously been observed in equity markets (see e.g. [77] and [93]) and derivative markets ([66]). Microstructure literature offers four conjectures as possible reasons for price clustering. According to the price resolution hypothesis proposed by [14], traders resist quoting at higher resolution when facing greater uncertainty about an asset's fundamental value. The negotiation hypothesis of [85] posits that a small set of prices eases the negotiation process by precluding frivolous offers and counter-offers. The attraction hypothesis of [64] states that traders quote-cluster at certain prices due to their specific preferences. Finally, the collusion hypothesis proposes that dealers may collude to quote at larger price increments in order to get larger profits ([37]). Price resolution, negotiation and collusion hypotheses can be immediately rejected in our context. The attraction hypothesis is similar in spirit to what we observe. On the price clustering aspect of our analysis, this paper is different from previous studies in that we show the price clustering preference of a sub-population of market makers–the manual traders–is revealed by tick size change.

3.3. Interdealer FX Market

While algorithmic trading has become a pre-requisite for market making in equity markets, manual traders still play an important role in the interdealer foreign exchange market, where large banks and institutions trade with each other in units of million, as liquidity suppliers.[6] Also in contrast to the highly fragmented equity market, the FX market practice of trading via "vehicle currencies" leads to liquidity concentration not just in trading venues but a handful of currency pairs.[7] Trading in the interdealer FX spot market is dominated by two venues. Reuters Matching is the primary trading venue for commonwealth and emerging market currencies and EBS is the leading liquidity provider for currency pairs EUR/USD, USD/JPY, EUR/JPY, USD/CHF and EUR/CHF.[8]

Given the monopoly of the EBS platform in the interdealer market on major currencies, manual traders–whom we observe to be forced to concede trading options–did not relocate their trading after tick size change. Given the structure of the market, an individual manual trader has no incentive to transfer his liquidity elsewhere alone and there is positive network externality only if all manual traders coordinate to move their trading activity *en masse*. It is difficult to implement such a coordinated move given the current structure of the interdealer market.[9] On the contrary, limit order book evidence points to manual traders remaining in the market after tick size change.

The causal relationship between the changes we observe and intensified algorithmic trading behavior is based on exogeneity of tick size change in the FX institutional environment. First, we address the possibility that tick size change is correlated with macroeconomic factors. Specifically, as the primary source of information impacting the FX market is significant macroeconomic events, of possible concern to us is the possibility that tick size change coincided with a reduction in intensity or number of economic informational events. It is unlikely that reasons behind a microstructure change is correlated with the macroeconomy. Furthermore, significant economic events that impact the FX market are directly visible via exchange rate volatility and we do not observe any changes in volatility patterns–see Figure 3.5, Figure 3.6, and Table 3.1 before and after tick size change. On the microstructure level, it is unlikely that anticipation effect drives the changes we observe in trader behavior that are relevant to our results. The potential quote-placement

[6] For example, the Swiss National Bank was an active manual trader on the EBS platform during the period we analyze.

[7] For example, a trader converting Canadian dollar to Mexican Peso will probably first trade Canadian dollar for a more liquid vehicle currency such as the US dollar then trade US dollar for Peso. The two dominant currencies are the US dollar and Euro. The EUR/USD currency pair accounts for 28% of global FX turnover. Individually the US dollar and Euro are involved in approximately 75% and 46% of all spot transactions respectively. See, for example, the BIS report: `http://www.bis.org/publ/rpfx10.pdf`.

[8] FX market convention is to list base currency first. For example, EUR/USD should be read as "US dollar per Euro".

[9] The proliferation of algorithmic trading on EBS and Reuters led a group of leading FX banks to consider the possibility of forming a bank-only trading platform called FXPure in 2011. This project was eventually abandoned due to lack of interest.

strategies of both manual and algorithmic traders arguably remained the same after tick size change. In fact for manual traders, their actual quote-placements did not change. For trading algorithms, cost in adjusting to decimal pip is likely to be low. Jumping in front of static orders may already be part of existing repertoire of a trading algorithm but binding spread prior to decimalization prevented execution of such tactic.[10] More importantly, it is highly unlikely that algorithmic traders were anticipating trading options made available by manual trader behavior–manual traders would have to alert algorithmic traders in advance of their concession of trading options.[11]

3.4. Description of Data

The data used in this study is the EBS limit order book at highest resolution available, which includes 10 levels of quotes on both the bid and ask sides at 100 milliseconds frequency for EUR/USD currency pair. This is the same as tick-by-tick snapshots of limit order book seen by traders. The deal time is rounded to the nearest 100 millisecond and only best buyer or seller initiated transactions are reported. Orders in EBS must be submitted in units of millions of the base currency. The period we analyze is the entire two year period from January 2010 to December 2011. There are approximately 500 million snapshots of the EUR/USD order book and approximately 18 million recorded deals. FX markets trade continuously and each trading day in EBS is 24 hours beginning and ending at 17:00 US Eastern Standard Time (21:00 Greenwich Mean Time). We exclude thin weekend trading periods and holidays as the liquidity tend to be extremely limited during these periods. The time stamps in the data are in GMT which varies due to daylight savings. We control for daylight savings time and standard time.[12] We present our results on the EUR/USD currency pair.[13]

3.5. Jumps

3.5.1. Semi-martingale Model of Price

According to asset pricing theory, an arbitrage-free price process must be a semi-martingale. Following [6] and [4], for econometric tractability we assume the prices, currency exchange rates in our case, follows an Itô-semimartingale. In this section, we summarize the model of price and econometric methodology.

[10] In our sample, the binding probability of spread was approximately 60% during the 14 months before tick size change and 1% during the 10 months after.

[11] Not only is there no anticipation effect, changes in adverse selection measures as well as the pattern of trading behavior between manual and algorithmic traders we describe below persist throughout our sample—10 months after tick size change was instituted.

[12] Similar conventions were adopted by [9] and [33].

[13] Results on the other four major currency pairs do not differ qualitatively from EUR/USD.

An Itô semi-martingale is a stochastic process of the form[14]

$$y(t) = \alpha + \int_0^t \sigma(s)dw(s) + \sum_{i=1}^{N(t)} j_i$$

where the summand processes are independent and

1 α is a process for which almost all paths are continuous and of finite variation.

2 w is standard Brownian motion.

3 The Itô integrand σ, the *spot volatility process*, is pathwise strictly positive, cádlág, and locally bounded away from zero.

4 $N(t)$ is a finite activity, simple counting process with, for all $t > 0$, $N(t) < \infty$ almost surely and $\{j_i\}$ is a countable family of non-zero random variables.[15]

5 (α, σ) is independent of w.

Assumptions 1 and 2 describe general Itô-semimartingales. The strict positivity of the spot volatility process can be assumed per Assumption 3 because thin weekends and holidays are excluded from our data. Assumption 4 specifies that, with probability 1, sample paths of the price process have finitely many jumps on $[0, t]$ for all $0 < t < \infty$.[16] This assumption is empirically justified by that average time-between-trades in our data is approximately 2.5 seconds. Assumption 5 precludes leverage effect, i.e. the negative correlation between volatility and returns. The FX market is subject to factors such as central bank interventions that make the existence of leverage effect not apparent.

3.5.2. Estimation of Volatility and Jumps

Denote the jump component of $y(t)$, $\sum_{i=1}^{N(t)} j_i = y(t) - y(t_-)$, by $\Delta y(t)$. The *integrated variance* of $y(t)$ is $c(t) = \int_0^t \sigma^2(s)ds < 0$ for all $t < \infty$.[17] The *quadratic variation*, or *square bracket process*, of $y(t)$ is defined as

$$[y](t) = c(t) + \sum_{s \in [0,t]} (\Delta y(s))^2$$

and can be consistently estimated by *realized volatility*

$$[y](t) = \text{plim}_{M \to \infty} \sum_{j=1}^{M} (y(t_j) - y(t_{j-1}))^2$$

[14] Adaptedness with respect to an underlying filtration is assumed throughout this section. We suppress notation for readability. Similarly, stopping times are measurable with respect to the underlying filtration.

[15] $\{j_i\}$, and the inter-arrival times of $N(t)$ are therefore time series and analyzable using discrete-time techniques, as is done in 3.5.6.

[16] A semimartingale can have infinitely many jumps on a compact interval, e.g. an infinite activity Lévy process.

[17] By Itô isometry, $c(t) = Var(\int_0^t \sigma(s)dw(s))$.

where $t_0 = 0 < t_1 < \cdots < t_M = t$ are stopping times with $\lim_{M\to} \sup_{1\le j\le M} t_j - t_{j-1} \to 0$ almost surely.[18] We use the bipower variation technology of [16] for separating integrated volatility and jump contribution.[19] Define the notation

$$\mu_r = \mathbb{E}[|u|^r] = 2^{\frac{r}{2}} \frac{\Gamma(\frac{1}{2}(r+1))}{\Gamma(\frac{1}{2})}, \text{ where } u \sim \mathcal{N}(0,1)$$

where $u \sim \mathcal{N}(0,1)$ and Γ denotes the Gamma function. For $r \in (0,2)$,

$$\frac{1}{\mu_r \mu_{2-r}} \{y_M\}_i^{[r,2-r]} \xrightarrow{p} \int_{h(i-1)}^{hi} \sigma^2(u) du$$

The difference between realized volatility and bipower variation can therefore be used to detect jumps. According to [16], under the null that the sample paths have no jumps,

$$\frac{\log(\sum_{j=1}^{[\frac{t}{\delta}]-1} y_j^2) - \log(\frac{1}{\mu_1^2} \sum_{j=1}^{[\frac{t}{\delta}]-1} |y_j||y_{j+1}|)}{(0.6091 \cdot \max\{\frac{\delta}{t}, \frac{\sum_{j=1}^{[\frac{t}{\delta}]-3} |y_j||y_{j+1}||y_{j+2}||y_{j+3}|}{(\sum_{j=1}^{[\frac{t}{\delta}]-1} |y_j||y_{j+1}|)^2}\})^{\frac{1}{2}}}$$

converges in law to $\mathcal{N}(0,1)$.

3.5.3. Microstructure Noise

To choose a sampling frequency, we compare the difference between realized volatility $[y]$ and bipower variation $\frac{1}{\mu_1^2}\{y_M\}_i^{[1,1]}$ across different frequencies. In the absence of microstructure noise, the difference consistently estimates the quadratic variation of the jump component:

$$\{y_M\}_i^{[2]} - \frac{1}{\mu_r \mu_{2-r}} \{y_M\}_i^{[r,2-r]} \xrightarrow{p} \sum_{k=N(h(i-1))+1}^{N(h(i))} j_k^2$$

The stabilization of the difference with respect to frequency therefore indicate absence of microstructure noise. For all five major currency pairs in the EBS market, hourly averages over the entire year of 2011 are computed for both realized volatility and bipower variation. It is interesting to observe that the difference $\{y_M\}_i^{[2]} - \frac{1}{\mu_r \mu_{2-r}} \{y_M\}_i^{[r,2-r]}$ stabilizes around the same frequency, every 30 seconds, for all five pairs. Figure 3.1 shows the results for EUR/USD and Figure 3.2 for USD/JPY. The magnitude of the stabilized gap, however, decreases with respect to the degree of liquidity empirically observed in the market. The most liquid currency pair, EUR/USD has the smallest difference between realized volatility and bipower variation at the stabilized frequency of 30 seconds. In other words, empirically, microstructure noise disappears at the same frequency for all five major currency pairs. The

[18] In our specific case, we choose to sample at regular time intervals. So the stopping times are in fact deterministic.

[19] Informally, the bipower variation does not see the contribution of jump component because large jumps do not occur between two adjacent intervals as the intervals become sufficiently small.

more liquid a currency pair, the more of its variation is due to the continuous part of the exchange rate process, instead of the jump component.

Figure 3.1: RV - BP, EUR/USD

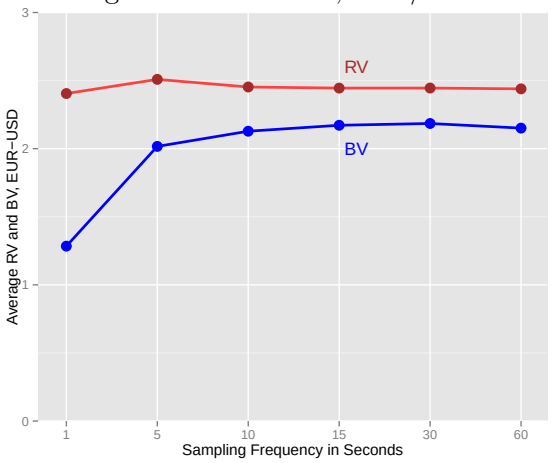
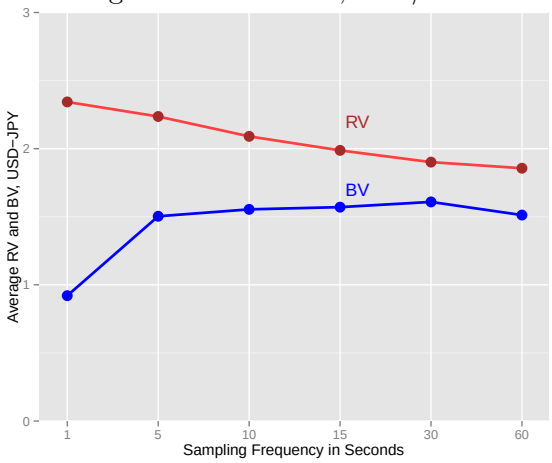


Figure 3.2: RV - BP, USD/JPY



3.5.4. Confirmation of Stylized Facts

En route to our results on serial correlation, we confirm two stylized facts regarding the foreign exchange market. Unlike the U-shaped volatility smile of (US) equity market, the FX market intraday volatility typically is hump-shaped with two peaks (see also [57], Figures 6.1 and 6.2). While trading can be done at any time, peak volatility occur at the opening times of US and European markets. There may be a smaller third peak that coincides with the opening of Asian markets. Two sample days, one before and one after tick size change, are shown in Figures 3.3 and 3.4. Furthermore, extreme high volatility –orders of magnitude larger than that of the typical intraday seasonality– tend to coincide with macroeconomic news releases or otherwise significant economic events. We show some representative instances of such news-driven high volatility periods for EUR/USD currency pair. Figures 3.5 and 3.6 show the estimated hourly integrated volatility of EUR/USD of 2010 and 2011 respectively. The integrated volatility time series show clearly visible peaks, the eight highest of which are circled in red. All eight periods of extreme high volatility coincide with significant economic news or events. Table 3.1 lists them below in chronological order, along with coincident economic events. The hourly periods of estimated extreme high volatility are given in Greenwich Mean Time (GMT), as is customary in FX market.

Table 3.1: EUR/USD Extreme Volatility Periods

Date	GMT	Corresponding economic Event
2010/05/06	18:00-19:00	The Flash Crash.
2010/05/19	14:00-15:00	Germany surprises the market by unilaterally banning short-selling against stocks and bonds, including sovereign bonds.
2010/05/20	17:00-18:00	European stocks plunge amid split over response to sovereign debt crisis.
2010/11/03	18:00-19:00	The Federal Reserve announces major quantitative easing plan to buy $600 billion in long-term treasuries over the next eight months.
2011/06/29	12:00-13:00	The Greek parliament passes new austerity package measures amid unrest.
2011/08/09	18:00-19:00	The Federal Reserve announces it intends to keep interest rate at exceptionally low levels –between 0 to 0.25 percent– through mid-2013.
2011/09/06	08:00-09:00	The Swiss National Bank announces decision to no longer tolerate EUR/CHF below CHF 1.2.
2011/11/30	13:00-14:00	The Federal Reserve and central banks around the world announce joint policy to alleviate the Eurozone crisis.

All macroeconomic news announcements are captured on the hour. For example, the Federal Reserve announcement regarding the Fed Funds rate are consistently made within a few minutes of 2:15pm US Eastern Standard Time (EST) of FOMC meeting day, which is 18:15 GMT. The joint announcement on 2011/11/30 by the Federal Reserve, Bank of Canada, Bank of England, Bank of Japan, European Central Bank, and Swiss National Bank occurred on 8:00am EST, which is 13:00 GMT as shown in the table.[20] Similarly, the Flash Crash occurred on 2:45pm EST, which is 18:45 GMT. We point out that, unlike equity markets, where price discovery is driven within-firm factors such as corporate governance and idiosyncratic innovations as well as industry wide shocks, exchange rate dynamic in the FX market reflects informational events such as central bank announcements.

[20] See, for example, `http://www.federalreserve.gov/newsevents/press/monetary/20111130a.htm` and `https://www.ecb.europa.eu/press/pr/date/2011/html/pr111130.en.html` for corresponding press releases of the Federal Reserve and European Central Bank respectively.

Figure 3.3: Intraday Seasonality of EUR/USD, 2010

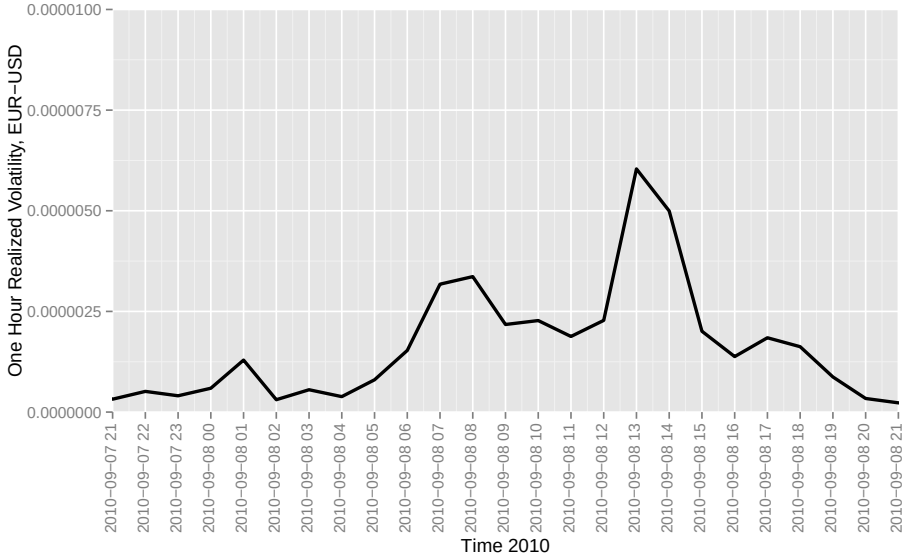
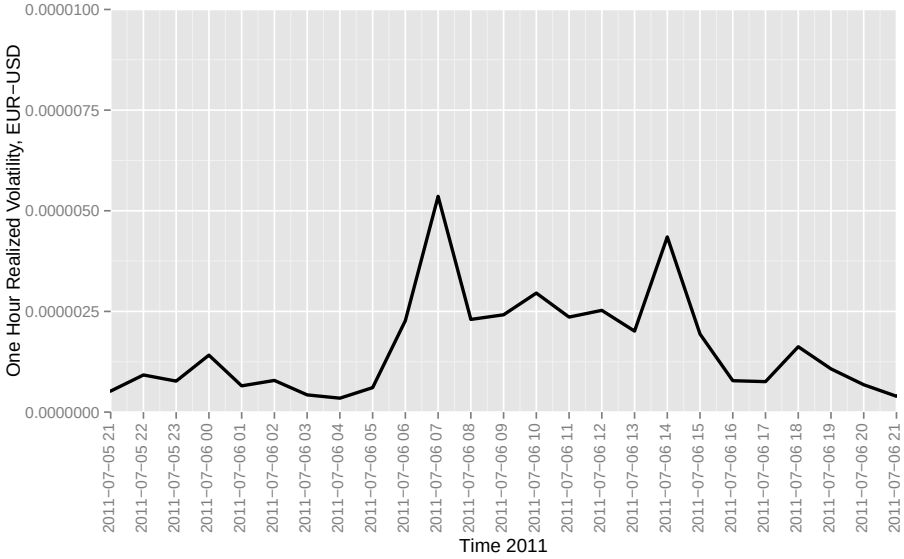


Figure 3.4: Intraday Seasonality of EUR/USD, 2011



*Notes:*Intraday seasonality of EUR/USD volatility before and after tick size change.

Figure 3.5: Integrated Volatility of EUR/USD, 2010

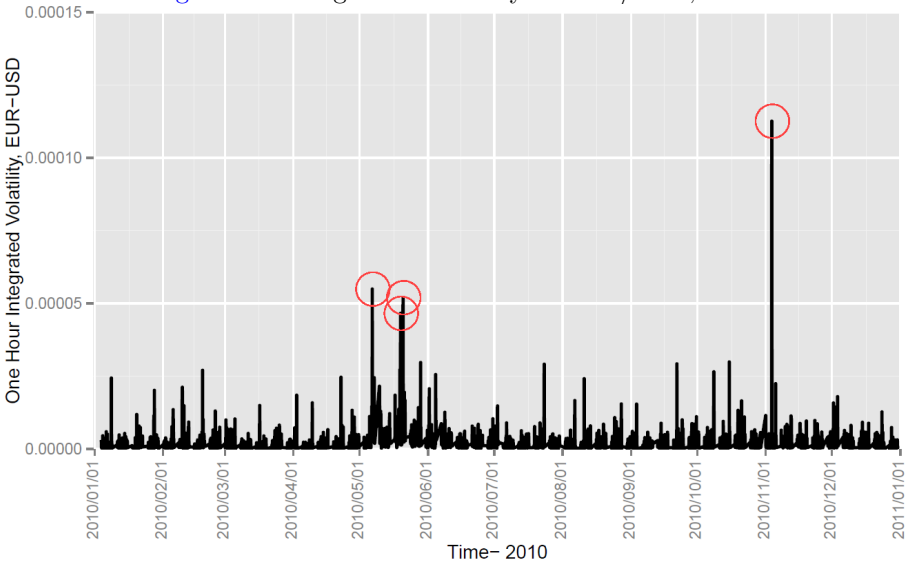
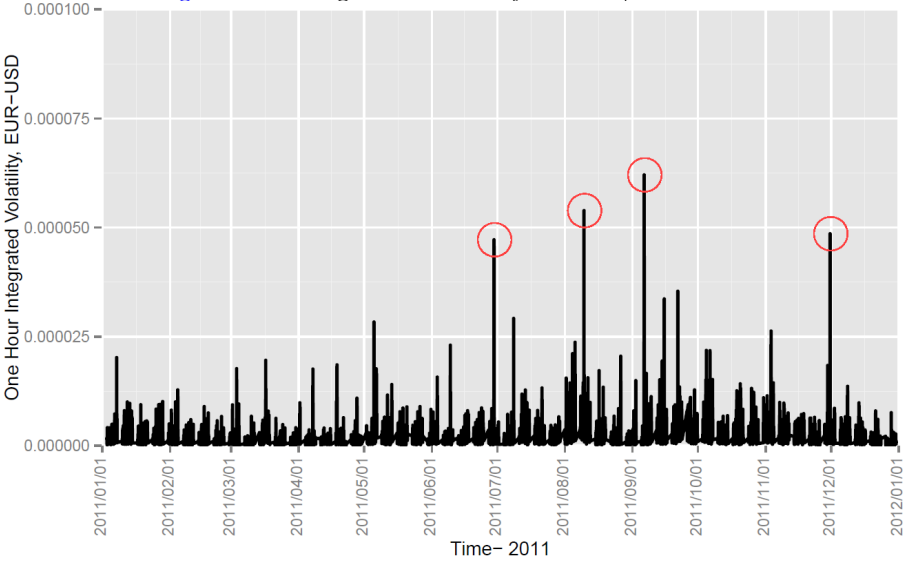


Figure 3.6: Integrated Volatility of EUR/USD 2011



Notes: 2010-2011 estimates of hourly integrated volatility of EUR/USD before and after tick size change.

3.5.5. Jumps Before and After Tick Size Change

We first show the empirical findings from applying the jump estimation methodology outlined in Section 3.5.1. For each currency pair, we estimate the size and timing of the jumps for the two years 2010 to 2011 at the frequency of every 30 seconds. In other words, given the observed sample path of $y(t)$, we estimate the realization of the jump time series $\{j_i\}$ (indexed by random arrival times of the counting process $N(t)$). Figures 3.7 and 3.8 show the histogram of estimated realization of $\{j_i\}$ for EUR/USD, i.e. the empirical distribution of jump sizes before and after tick size change, respectively. Not surprisingly, introduction of decimal pip cuts out a neighborhood around zero from the distribution of jump sizes. Figures 3.11 and 3.10 show the histograms of estimated inter-arrival times before and after tick size change. Both can be reasonably fitted by exponential distributions. Figures 3.11 and fig:Daily-Jumps-After shows the daily time series of number of jumps, spanning the years 2010-2011.

3.5.6. Serial Correlation Properties of Jumps

The serial correlation properties of the jump component Δy of the currency pair EUR/USD are reflected in the serial correlation properties of the jump time series $\{j_i\}$ and inter-arrival times of jumps. We test $\{j_i\}$ for serial correlation and fit appropriate time series models to the inter-arrival times. The data generating process of jumps undergoes unambiguous structural change in serial correlation properties before and after tick size change. The reduction (complete dissipation, in the case of $\{j_i\}$) of serial correlation in both jump time series and inter-arrival times means the jump component $\sum_{s\in[0,t]} \Delta y(t)$ of exchange rate becomes significantly more martingale-like after tick size change.[21]

Inter-arrival times Inter-arrival times exhibit autocorrelation both before and after tick size change. Fitting auto-regressive models to the before and after inter-arrive times series yields AR(27) and AR(8) models, respectively. Therefore the autoregressive lag decreases sharply across tick size change. Diagnostic tests of residuals, shown in Figures 3.13 and

[21] For example, jump sizes and inter-arrival times of the compound Poisson process –a basic building block of Lévy processes– are both white noise time series.

3.14, confirm model specification of the autoregressive model. The arrivals of jumps are significantly less clustered after tick size change.

Figure 3.7: Jump Sizes of EUR/USD, Before Tick Size Change

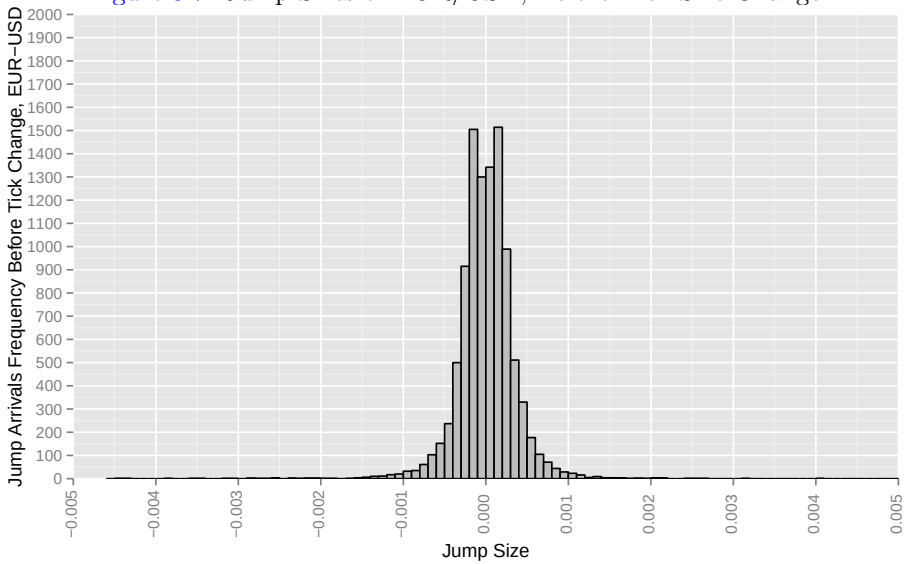
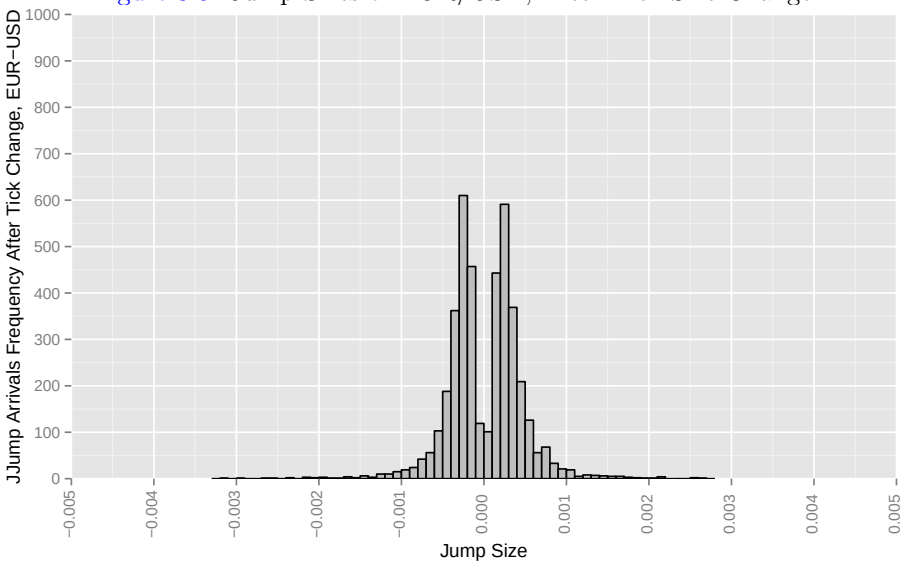


Figure 3.8: Jump Sizes of EUR/USD, After Tick Size Change



Notes: Figures 3.7 and 3.8 show the histograms of jump sizes before and after tick size change.

Figure 3.9: Jump Inter-Arrival Times Before Tick Size Change, EUR/USD

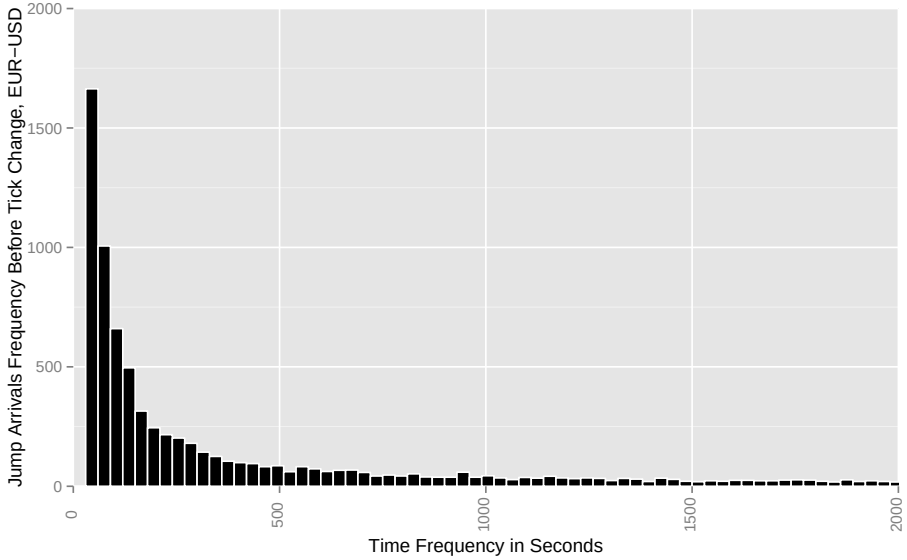
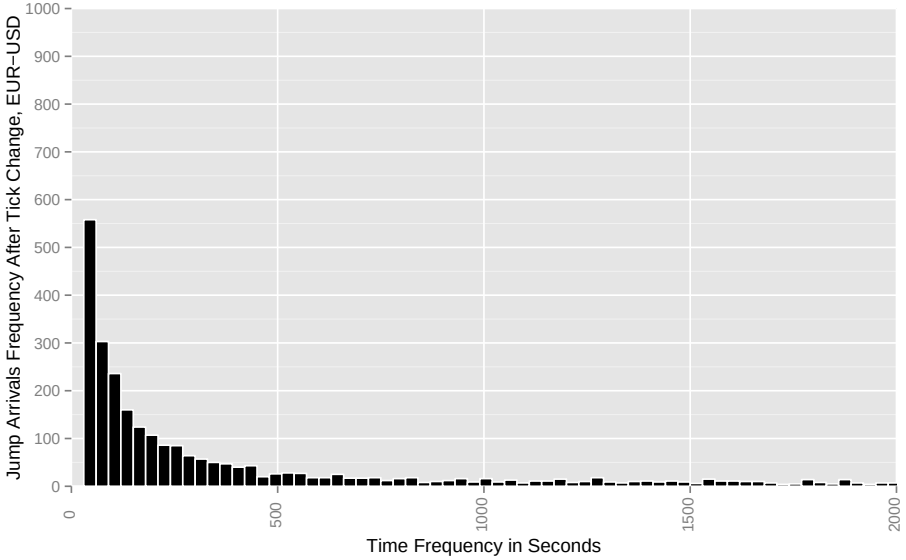


Figure 3.10: Jump Inter-Arrival Times After Tick Size Change, EUR/USD



Notes: Figures 3.9 and 3.10 provide the histograms of jump inter-arrival times before and after tick size change. Exponential distributions are fitted, with results shown in 3.2.

Figure 3.11: Daily Number of Jumps of EUR/USD, Before Tick Size Change

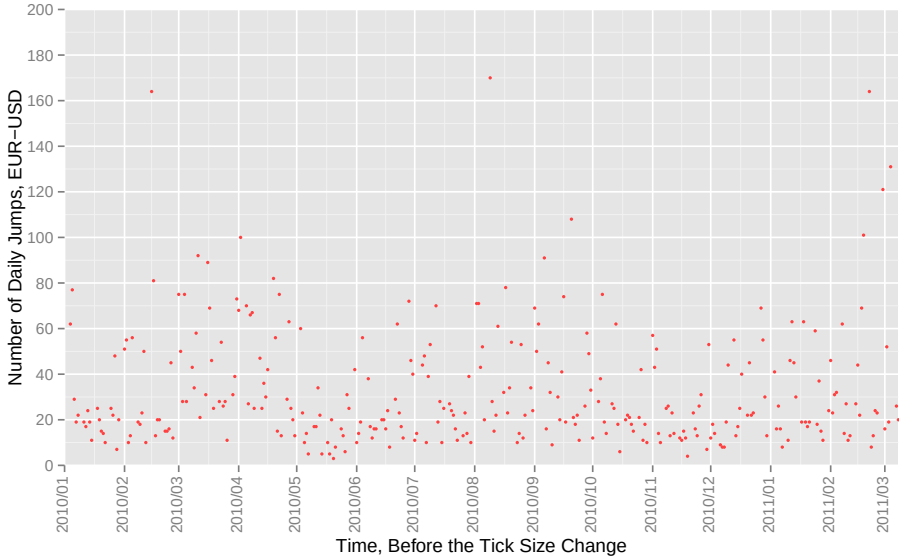
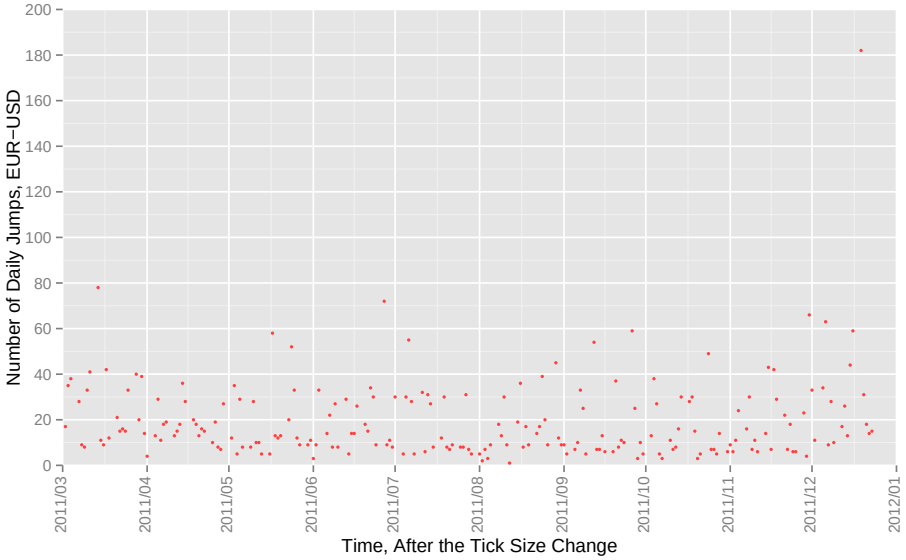


Figure 3.12: Daily Number of Jumps of EUR/USD, After Tick Size Change



Notes: Figures 3.11 and 3.12 illustrate the number of daily jumps before the tick size change respectively.

Figure 3.13: Jump Inter-Arrival Times, Before Tick Size Change

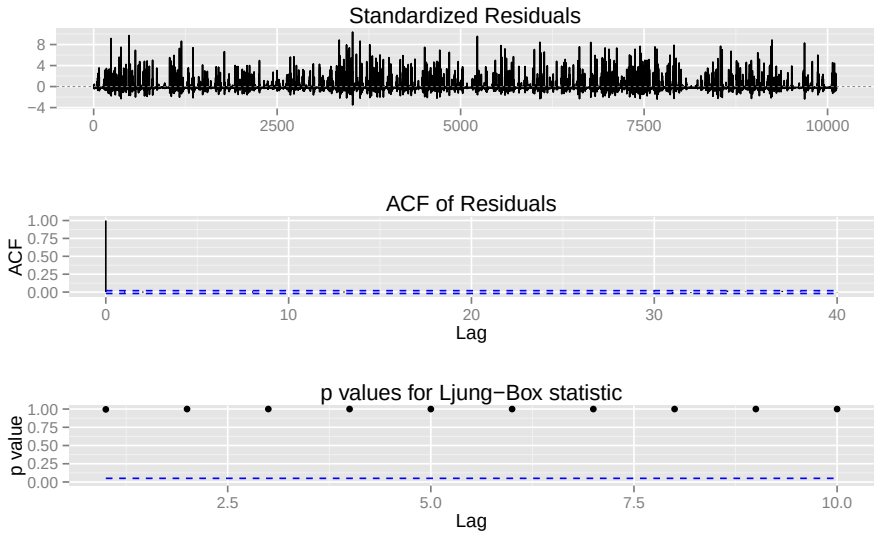
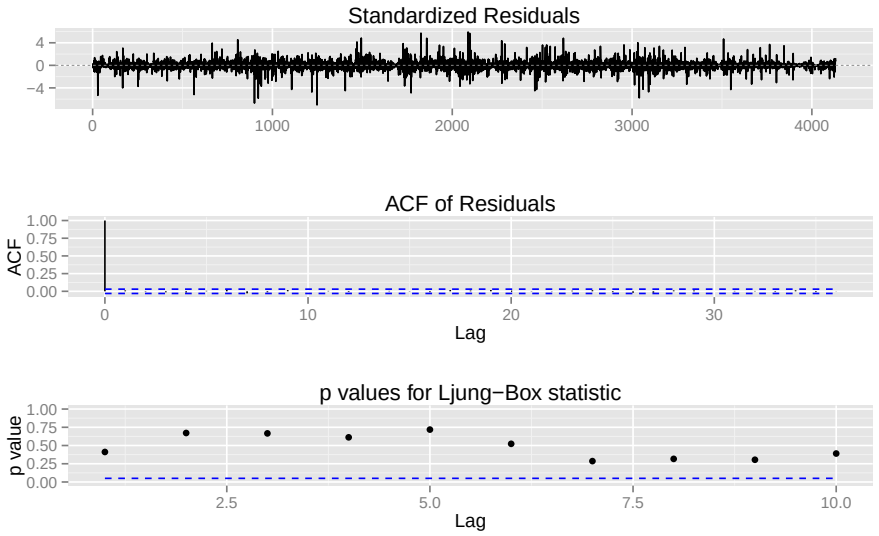


Figure 3.14: Jump Inter-Arrival Times, After Tick Size Change



Notes: Autoregressive time series fit and diagnostics for jump inter-arrival times. The Akaike Information Criterion achieves the minimum value of zero for the AR orders chosen.

Distribution of jump sizes Jump size distributions are centered near zero both before and after tick size change. The kurtosis of the distribution shows heavy tail both before and after tick size change. Nonparametric Mira test for symmetry rejects the unconditional jump size distributions before tick size change at $p < 2.2 \times 10^{-16}$ and does not reject the distributions after tick size change at 1% level of significance. Jumps sizes having a symmetric distribution centered at zero imply that, in the long run, large market orders have zero permanent price impact, implying no information content.

Jump time series Taking the time series dimension into account, the Box-Ljung serial correlation test rejects the jump sizes time series $\{j_i\}$ before tick size change at 1% significance level with $p = 0.005544$. After tick size change, the jump sizes time series has $p = 0.8185$. The time series sample sizes are 10,150 and 4,214 before and after respectively. The lag chosen for the Box-Ljung test is 40.[22] Results are summarized in Table 3.2. A jump process driven by market taker's private information necessarily has serial correlation both in direction and arrival times. Less clustering of jump arrivals, increased symmetry and disappearance of serial correlation in jump sizes all point to a reduction in information content in jumps, in both time and direction dimensions. We can conclude that for the EUR/USD currency pair, large liquidity consumption after tick size change is much less likely to be driven by market taker's private information than before.[23]

3.6. Components of Quoted Spread

Jumps in currency prices only directly captures large liquidity consumption activity. To provide further evidence of curtailment of information asymmetry across tick size change, we examine the limit order book using the model of [76]. The model posits that

$$\Delta P_t = \delta + \frac{S}{2} Q_t + (\alpha + \beta - 1)\frac{S}{2} Q_{t-1} - \alpha(1 - 2\pi)\frac{S}{2} Q_{t-2} + \epsilon_t$$

[22] The less powerful non-portmanteau Box-Pierce and Durbin-Watson tests are also performed. Although they are unable to reject, both p-values increase after tick size change.

[23] While serial correlation of large order flow may also be due to traders splitting up larger orders, it not clear that occurrence of order splitting in the interdealer FX market is as common as in equity market. For example, EBS offered no iceberg order type—a common microstructure product on equity trading venues—during the two year period we analyze, possibly due to lack of demand. In any case, the main driver that we argue to be behind the drastic reduction in serial correlation— manual market makers being forced to take the market after tick size change—is orthogonal to order splitting considerations.

Table 3.2: Jump Component of EUR/USD Process Before and After Tick Size Change

Inter-arrival times	Before	After
Daily average number of jumps	33.28105	19.60094
Arrival intensity (exponential fit)	0.0003982155	0.0002353829
Autoregressive order	27	8
Jump sizes	**Before**	**After**
Mean	6.286406×10^{-6}	1.059269×10^{-5}
Standard deviation	0.000368071	0.0004490323
Skewness	1.191105	-0.1275948
Kurtosis	67.60528	4.543458
Mira symmetry test	$p < 2.2 \times 10^{-16}$	$p = 0.01928$
Box-Ljung Test	$p = 0.005544$	$p = 0.8185$
Box-Pierce Test	$p = 0.3558$	$p = 0.468$
Durbin-Watson Test	$p = 0.1767$	$p = 0.2297$

where P_t is the transaction price, S is the traded spread, Q_t is the trade indicator process, 1 if buyer initiated and -1 if seller initiated. The parameter α is the portion of S due to adverse selection, β is the portion of S due to inventory risk, and π is the probability of a trade flow reversal. The trade indicator process is assumed to follow a Markov process

$$E[Q_{t-1}|Q_{t-2}] = (1 - 2\pi)Q_{t-2}$$

The error term ϵ_t contains both public information and the difference between traded spread S and the quoted spread. The latter may include, for instance, rounding error. The model is estimated using generalized method of moments (GMM). To enter a frequency where microstructure effects are explicitly present, we choose to aggregate the data every 5 seconds. Table 3.3 shows the results for EUR/USD. The traded spread before the tick size change is estimated to be approximately 6.4 decimal pips, while the minimum possible quoted spread is 1 pip. This agrees with the fact that there was binding pressure on the spread prior to tick size change. Before tick size change, the quoted spread was binding approximately 60% of the time.[24] The estimates of α and β sum to greater than 1 is also likely due to pressure on quoted spread. The average quoted spread is approximately twice that of estimated spread. Scaling the estimates of α, and β, before tick size change by

[24]In contrast, the binding probability of quote spread after tick size change is 1%.

0.5 still shows a significant reduction in adverse selection component of spread across tick size change. After tick size change the estimated traded spread is around 8.5 decimal pips, which is very close to the average quoted spread.

Table 3.3: Estimated EUR/USD components of spread, 5-seconds, Sample size = 163,069.

Components of the Spread	Before	After
α	0.85	0.17
β	1.05	0.23
π	0.21	0.31
S	0.00006	0.00008
δ	0.00014	-0.00059

All estimates are significant at 1% level of significance.

The estimated adverse selection component α decreases across tick size change. Estimates of the other parameters are also of interest. The inventory risk component β also decreases across tick size change. This is as we expect, since decimalization means market maker suffers smaller loss in disposing potentially toxic order flow. The probability of trade flow reversal π increases after tick size change. Therefore trade direction becomes less persistent not just for large liquidity consumers, as shown in Section 3.5.6, but for all liquidity consumers. The change in α and π are consistent with our hypothesis that tick size change attenuates the market maker's adverse selection problem.

3.7. Realized Spread

In addition to econometric evidence, we consider the empirical measure of realized spread of the order book before and after tick size change. Realized spread at time t, denoted by RS_t, is defined:

$$RS_t = 2Q_t(P_t - M_{t+s})$$

where P_t is the transaction price at time t, Q_t is the trade direction indicator, and M_{t+s} is the midpoint at time $t+s$ for some chosen time interval s.[25] RS_t is the difference between current deal price and the quoted midpoint at a future time. After a transaction at time

[25] Customary choice of lag s of realized spread is 5 minutes (e.g. [71]). This is not appropriate in our high-frequency setting.

t, price movements favorable to the market maker from t to $t+s$ results in positive RS_t, and vice versa.[26] According to our analysis in Section 3.5.3, microstructure effect ceases to be present at frequency lower than every 30 seconds. Our computation shows that the realized spread exhibits the same behavior across tick size change at all frequencies higher than every 30 seconds, that is, under different degrees of microstructure effect.

Figures 3.15 and 3.17 show the daily average realized spread of EUR/USD currency pair for 2011, before and after tick size change at frequencies of every 5 and 10 seconds. There is a clear positive shift in realized spread across tick size change. For EUR/USD, price movement for the market marker tend to be unfavorable with negative realized spread before tick size change, while realized spread is nearly zero after tick size change. The effective spread at time t is defined by $ES_t = |P_t - M_t|$. ES_t measures the revenue of the market maker from supplying immediacy. As a proxy for adverse selection of market making, we use the limit order book statistic

$$ES_t - \frac{RS_t}{2}$$

i.e. revenue from supplying immediacy minus loss due to adverse price move. Similar to that of realized spread, the behavior of the adverse selection proxy is consistent with respect to different levels of microstructure effect. Figures 3.16 and 3.18 show a clear downward shift for EUR/USD across tick size change at frequencies of every 5 and 10 seconds.

[26] Realized spread was shown to not reflect informed order flow by [40] in a specific equity setting where the informed trader is a corporate raider with long-lasting private information on how she may create value by influencing corporate governance. The information governing the FX exchange rates is fundamentally different—with the source of information being mostly macroeconomic events, as implied by the analysis in Section 3.5.4—and there is no analogous scenario in our setting. In our high-frequency setting, we take the restrictive definition that information is that regarding order flow. Realized spread increases after tick size change because it creates additional uninformed order flow from former market makers.

Figure 3.15: EUR/USD Realized Spread, Five Second Frequency

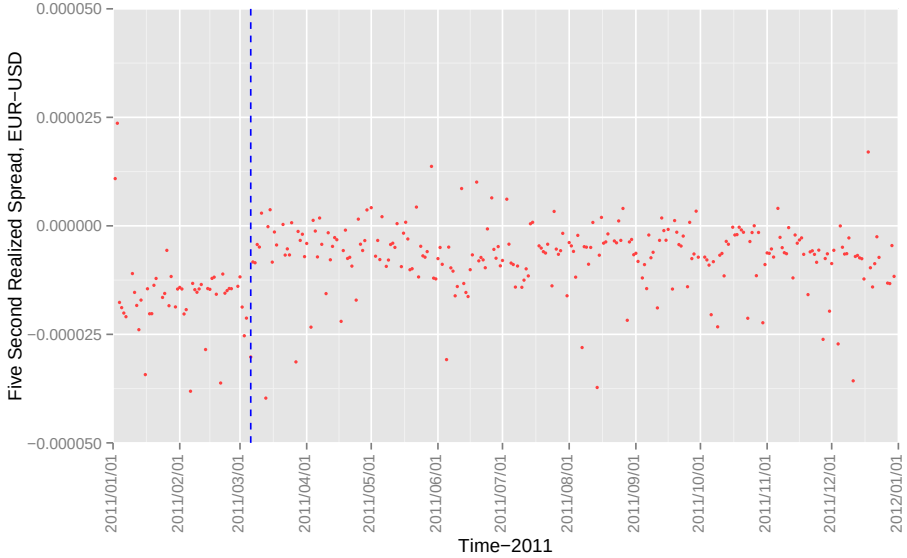
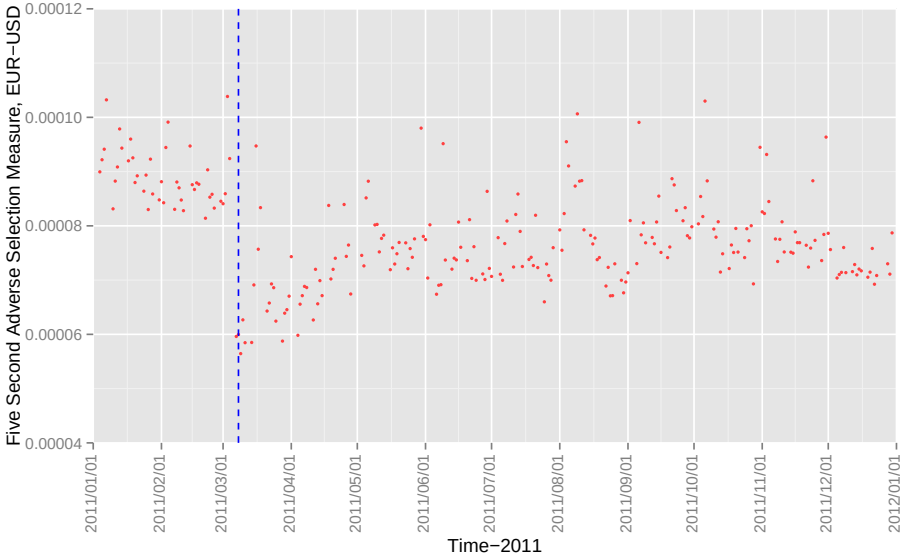


Figure 3.16: EUR/USD Adverse Selection Proxy, Five Second Frequency



Notes: Realized spread is measured at 5 seconds lag in Figure 3.15 and computed using 1.3 million deals before the tick size change and 7.3 million deals after the tick size change. Time series of daily averages are plotted. Tick size change is demarcated by blue line. Adverse selection proxy before and after tick size change in Figure 3.16 shows clear downward level shift.

Figure 3.17: EUR/USD Realized Spread, Ten Second Frequency

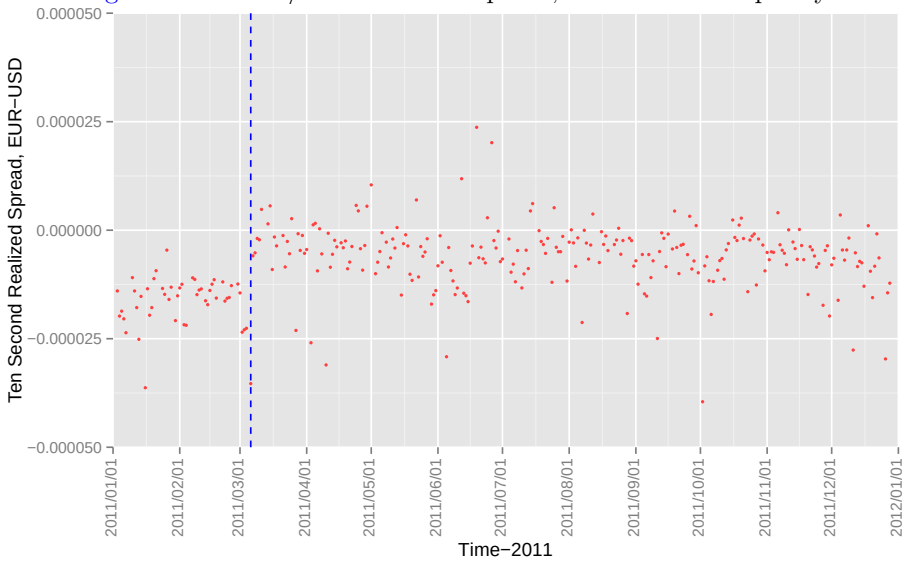
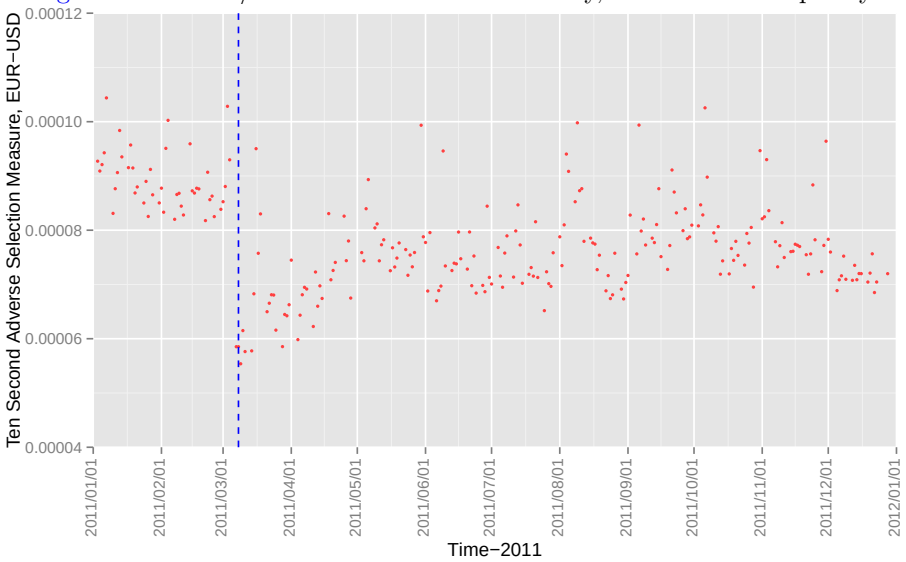


Figure 3.18: EUR/USD Adverse Selection Proxy, Ten Second Frequency



Notes: Realized spread is measured at 10 seconds lag in Figure 3.17 and computed using 1.3 million deals before the tick size change and 7.3 million deals after the tick size change. Time series of daily averages are plotted. Tick size change is demarcated by blue line. Adverse selection proxy before and after tick size change in Figure 3.18 shows clear downward level shift.

3.8. Market Participant Behavior

First we fix terminology by using the classification of traders in FX market given in the BIS 2011 report.[27] Market participants are divided into two major categories, Manual Traders and Automated Traders. Manual Traders use proprietary EBS workstations, for manual order management.[28] Automated Traders place orders algorithmically with little or no human intervention.[29] Automated Traders are capable of placing orders at a frequency far exceeding that for Manual Traders. Manual and automated market makers make up two distinct species of liquidity providers on the EBS platform. Manual market makers place limit orders for inventory or liquidity reasons, whereas their automated counterparts engage in opportunistic market making. Perhaps surprisingly, in sharp contrast with equity markets, manual trader presence dominates the interbank FX market. Using EBS client identity data, [103] makes the following observations:

- 75% of all traders in the EUR/USD pair are manual traders in 2011.
- The orders of manual market makers are filled in about 50% of the time before cancellation. In contrast, algorithmic market makers cancel 93% of their quotes.
- Manual market makers place large limit orders while automated market makers tend to submit orders of the minimum size one million. In fact, all orders larger than 4 million are from the manual market makers.

The clear reduction in adverse selection we observed in Sections 3.5.6, 3.6, and 3.7 can be explained by market participant behavior before and after tick size change. To infer market participant behavior, we now undertake an analysis on the evolution of shape of the limit order book. While we do not possess trader identities, a snapshot-to-snapshot inspection of the book shows clearly the behavior of automated market makers. Distributions of order

[27] See `http://www.bis.org/publ/mktc05.pdf` for more details. The same classification is adopted by EBS (see e.g. [103] and [32]).

[28] See `http://www.ebs.com/access-methods/ebs-workstation.aspx` for details on EBS workstations provided to Manual Traders.

[29] See `http://www.ebs.com/access-methods/ebs-ai.aspx` for details on EBS interface technology for automated trading.

sizes and quote price placements obtained from our anonymous data set are both consistent with snapshot-to-snapshot activity of automated traders and known characteristics of traders cited above. The tick size change revealed distinct preferences of the two species of market makers. Our results indicate that, while automated market makers engage in queue jumping after tick size change, manual market makers did not make use of newly available decimal prices in placing quotes.

Tick size change made one additional decimal place, the fifth, available to the market maker. The best bid and ask prices are predominantly concentrated at the old pip pricing levels after tick size change. Before tick size change, the last digits of the best digit prices are distributed uniformly as shown in Figures 3.19 and fig:Digit-Bid-After. The clear uniform distribution shows that all market makers make equal use of available prices in placing quotes. The distribution of last digits undergoes a clear, and somewhat surprising change after tick size change. After tick size change, the last digits of best limit prices are concentrated at 0, around 30% for both the best ask and the best bid as shown in Figures 3.21 and 3.22. While we have summarized the distribution of all snaps shots of the limit order book in our data before and after tick size change respectively, the two distinct distributions are stable at the daily level. This points to a significant portion of market makers who did not adapt to decimal pip pricing.

Our analysis suggests strongly that it is the manual traders who did not adapt. Table 3.4 shows the average order size at the best bid. Before tick size change, the average order size is uniformly distributed with respect to last digits, suggesting again that all traders make equal use of all available price levels when submitting quotes. After tick size change, average order size at prices with last digit zero, that is, at old pip pricing levels, is twice as large as those at the newly available decimal levels. In fact, orders placed at the newly available decimal pip levels have average size very close to the minimum order size of one million. Table 3.5 shows very similar results for the best ask. As automated market makers tend to submit orders of minimum size, this supports our claim that manual traders price-clustered at pip pricing –and, occasionally, half-pip pricing– levels.

Figure 3.19: Best Bid Last Digits, Before Tick Change

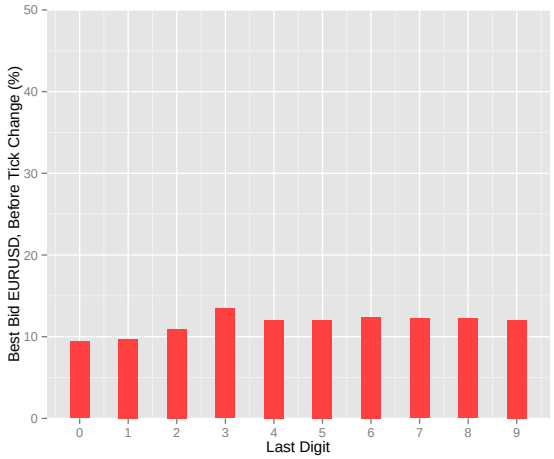


Figure 3.20: Best Bid Last Digits, After Tick Change

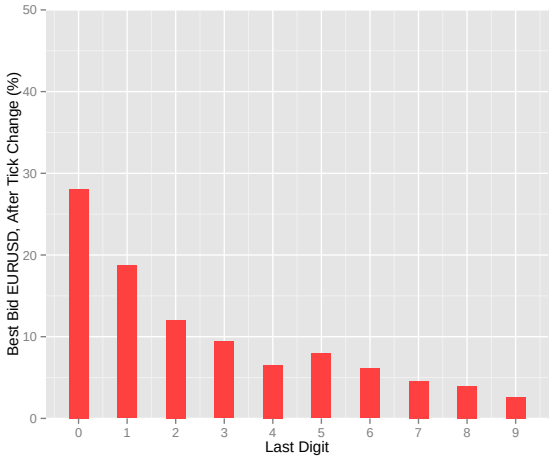


Figure 3.21: Best Ask Last Digits, Before Tick Change

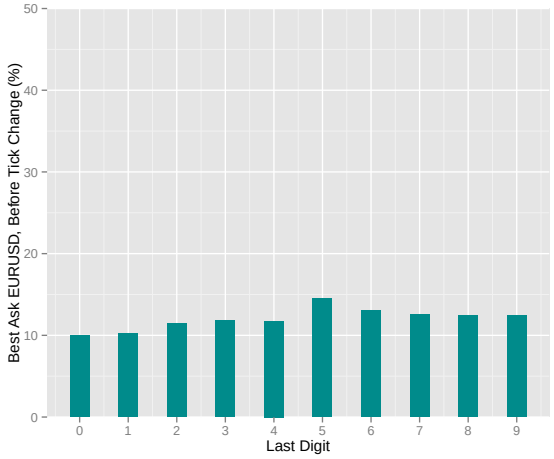


Figure 3.22: Best Ask Last Digits, After Tick Change

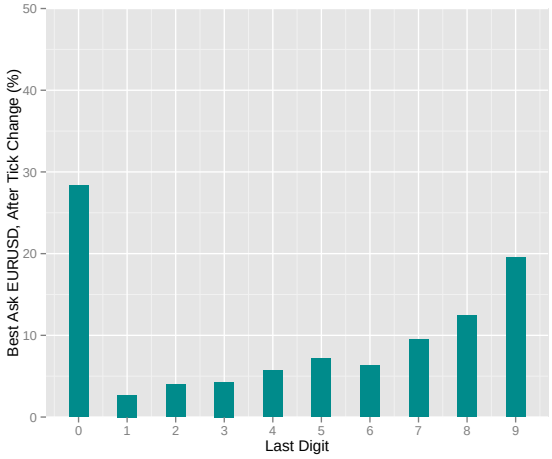


Table 3.4: Size of Average Order (Million) at the Best Bid

Last digit of quoted price	0	1	2	3	4	5	6	7	8	9
Before tick size change	1.59	1.43	1.57	1.52	1.54	1.61	1.54	1.60	1.50	1.41
After tick size change	2.12	1.06	1.05	1.08	1.08	1.48	1.15	1.11	1.07	1.10

Table 3.5: Size of Average Order (Million) at the Best Ask

Last digit of quoted price	0	1	2	3	4	5	6	7	8	9
Before tick size change	1.53	1.32	1.62	1.59	1.45	1.54	1.53	1.54	1.43	1.48
After tick size change	2.19	1.23	1.22	1.10	1.09	1.32	1.06	1.04	1.07	1.09

Automated market makers, on the other hand, have no reason to not take advantage of decimal pip pricing. Indeed, we show that automated market makers engages in queue

jumping. Queue jumping exploits the likely favorable price movement precipitated by a large static order, which also limits the loss for the queue jumper under unfavorable price movement. Static limit orders placed by manual traders at old pip pricing, with last digit zero, are vulnerable to queue jumping strategy by automated traders. On the ask side, such queue jumping is done by submitting an order with last digit 9 and the buy side by an order with last digit 1. Looking at ask side through the course of one trading hour, 36,000 snapshots of limit order book, at tick frequency, after tick size change shows that 43% of limit orders are at pip pricing levels (see Table 3.6). Conditional on the best ask being at old pip prices, the probability that the next snapshot shows one tick size improvement is 55%, with average order size around minimum order size. One or two tick size improvement makes up 70% of next-snapshot possibilities at the best ask, all with around average minimum order size. Same pattern is mirrored on the buy side of the book. As time between snapshots is 100 milliseconds, this directly exposes price clustering tendencies at pip prices—a behavior that can only be reasonably attributed to manual traders —and subsequent queue jumping activity by automated traders.

Table 3.6: Queue Jumping in One Trading Hour After Tick Size Change

Number of limit order book snapshots	36,000
Best ask at old pip pricing	15,490
One tick price improvement next snapshot	8,485 (avg order size 1.06)
Two tick price improvement next snapshot	2,289 (avg order size 1.04)
Best bid at old pip pricing	14,668
One tick price improvement next snapshot	7,767 (avg order size 1.03)
Two tick price improvement next snapshot	2,055 (avg order size 1.02)

The decreasing pattern shown in Figure 3.20 of last digit distribution of best ask price can be explained by the interaction we describe between traders. An initial static quote might be placed at a pip price by a manual traders. Automated traders queue-jump first the manual trader quote then leap frog each other as they compete for top of the book. Figure 3.22 tells the same story on the other side of the book. The already dramatic inequality in speed between manual and automated traders is accentuated by the tick size change. Clinging to old pip pricing makes limit orders submitted by manual market makers prey to

the queue jumping strategy. Decimalization of tick size therefore changes the information balance between market makers and takers in the market.

3.9. Summary

Analyzing both the price process and the limit order book, we showed that the degree of information asymmetry exhibited a discrete change across tick size change and provided an explanation on the relationship between these two events. A smaller tick size has two textbook counteracting effects on volume of informed order flow. Larger minimum tick means the spread is more likely to bracket the fundamental value, decimalization incentivizes the market taker to become more informed. On the other hand, as smaller tick size encourages predatory market making and makes informed trading less profitable, the market taker has less incentive to acquire information. However, the ecology of the interdealer FX markets, with no evidence of trader emigration or immigration across tick size change, gives rise to a different response driven by interaction between two distinct subspecies of market makers. Automated market makers crowd out their manual counterparts from the top of the book.

Current literature concerning the issue of the effect of algorithmic trading on market efficiency for the FX market have generally been favorable. For example, using EBS data from 2004 to 2008, [32] showed empirically that high frequency returns are serially uncorrelated and conclude that algorithmic trading activity improves market efficiency. There has also been some evidence that algorithmic trading contributes to the speeding up of the price discovery of exchange rates with respect to macroeconomic news (see e.g. [9] and [54]), which is partially corroborated by our analysis of periods of extreme volatility shown in Table 3.1.[30,31] To the best of our knowledge, this is the first paper that considers market efficiency and algorithmic trading from a microstructure perspective by analyzing comprehensively the impact of a specific microstructure event. In contrast to previous studies, our

[30] [32], [9] and [54] all use discrete time models and sample at lower frequencies.

[31] According to EBS itself: automated trading "...is a key component of the professional spot FX market place, offering efficient price discovery and 24-hour access to tight liquidity..."—`http://www.ebs.com/access-methods/ebs-ai.aspx`.

findings do not support the position that algorithmic trading improves market efficiency.[32] In our high frequency continuous-time setting, the price semi-martingale is the independent sum of a finite-variation process, a continuous martingale, and a jump process. In other words, the exchange rate has a locally riskless component, an informationally efficient component, and a jump component. While the unequivocal whitening of the jump component moves exchange rate closer to the martingale property that characterizes an informational efficient market ([50]), this is not a verification of market efficiency—despite observing uniform reduction in traditional measures of adverse selection in addition to whitening of jumps. This abatement of adverse selection is an artifact of market microstructure and algorithmic trader behavior, rather than genuine dissemination of information across market participants. Our study highlights that traditional microstructure metrics, though often indicative of the general market conditions, need to be complemented by analysis of market participant behavior to obtain a meaningful picture of the state of market, especially given the increasing pervasiveness of high-speed electronic trading.[33]

[32] Given the likely vast difference in preferences between a large institutional FX dealer and an algorithmic trader trading under prime brokerage, one may conclude that conceding the top of the limit order book is the optimal choice for manual traders optimizing against market microstructure and algorithmic trader behavior. However, EBS introduced an order type–"pip discretion rule"–in late 2011 that is explicitly designed to curb queue jumping. Conversations with traders suggest the introduction of pip discretion rule is due to manual trader complaints. See `http://www.ebs.com/~/media/Files/I/Icap-Ebs/infosheets/EBS-Market-info-sheet-Mar-2013.pdf`.

[33] Having called attention to the inadequacy of current financial econometric techniques in addressing the impact of algorithmic trading on market quality, we hesitate to venture an opinion on the general issue of social benefit of algorithmic trading.

Chapter 4

Network Risk Premium

4.1. Asset Pricing and Financial Networks

This chapter contributes to the literature that lie in the intersection of financial networks and asset pricing. There is a growing body of recent literature that considers financial networks and systemic risk. [12] suggest that the effect of relatively large, well-connected banks, on systemic stability scales more than proportionately with their size: the impact of their collapse is not only due to their connectivity, but also the resulting loss of confidence of other banks. [109] proposed that the analysis of interconnectivity and transmission of distress among financial institutions can be decomposed into bilateral interdependence linkages. [55] considers macro-prudential capital requirements in the context of a network banking model. [22] performs a principal components analysis on the returns of financial intermediaries to analyze systemic risk in the financial sector.

Also relevant is the analysis by [25] of the core-periphery structure in social networks. [74] offers an economic network from which a core-periphery structure arises and shows there is positive correlation between network centrality and payoffs. In this paper we propose a network pricing factor for asset returns that is derived from centrality.

Closest to our paper is [2] which uses network centrality as a pricing factor.[1] The difference between our paper and [2] is that, rather than using centrality in an off-the-shelf manner, we provide a microfoundation for our pricing factor and extend traditional factor

[1] See also [3] and [1].

pricing models, *a la* [51],[52] and [31], to a network setting. While [2] suggests that shocks might not cancel out through diversification but instead may aggregate into macroeconomic fluctuations, we argue through a corresponding stochastic discount factor that our network pricing factor reflects the diversification and exposure to contagion risk trade-off.

4.2. A Network Pricing Factor

We propose a network pricing factor which may reflect of contagion risk of adverse shocks from other firms in a financial network. A financial network of cross exposures can be represented as a weighted directed graph, with firms as vertices. We use the convention that incoming edges represent the exposure of a firm to other firms. That is, a directed edge from firm/vertex i to firm/vertex j indicates that firm j is exposed to firm i.[2] For a directed graph depicting a financial network of n firms, one can define an adjacency matrix $A = [a_{ij}]$, where a_{ij} is the amount of exposure of firm i to firm j.[3] Now a corresponding Markov Chain can be obtained by normalizing the directed edges outgoing from each firm so the edge weights sum to 1. In other words, the transition matrix T of this Markov chain is obtained by normalizing A column-wise so that each column sums to 1. The network pricing factor we propose f_ν is the stationary distribution of the Markov chain given by T. f_ν is characterized by the eigenvalue equation

$$T f_\nu = f_\nu.$$

and f_ν is has non-negative entries. It follows from the Perron-Frobenius theorem that f_ν is a Perron-Frobenius eigenvector of the matrix T (see, for example, [18]).[4]

The factor f_ν captures the likelihood that a firm in the network may be affected by an economic shock that originates within the network. A number of measures constructed using the Markov transition matrix T that purport to measure different types of network

[2] In this convention, the direction of an edge is the direction of possible travel of an adverse shock as it transmits through the network.
[3] In our notation, a_{ij} denotes the entry of A at i-th row and j-th column.
[4] Perron-Frobenius theorem is applicable subject to the technical condition that the corresponding graph is strongly connected.

flows are available. Remarks on the appropriateness of our choice in capturing contagion risk are in order. Other measures include degree, closeness, and betweenness. To choose the appropriate measure for the transmission of potential shocks in an economic network, one must consider the assumptions that underlie each measure. Network traffic could be assumed to follow a walk (an unrestricted sequence of nodes and links), a trail (a sequence in which no link is repeated), a path (a sequence in which no link or node is repeated), or a geodesic path (the shortest path between two nodes).[5] Furthermore, network traffic may spread serially (through only one path at a time), or in parallel (through multiple paths at the same time). Though making generalizations about the economic shocks is problematic, one can make a few reasonable assumptions about how shocks may flow from one firm to another. We consider the following two assumptions reasonable (see also [2]). First, regardless of how an economic shock is defined, it is unlikely to follow a geodesic path. Only traffic that has a known destination follows a geodesic path through the shortest distance (e.g., a courier delivering a package). In contrast, economic shocks that transmit across an economy do not have final recipients and are unlikely to follow the shortest path between firms/industries/economic sectors. This means that closeness and betweenness centrality are inappropriate for economic shocks since they implicitly assume that traffic follows geodesic paths.[6] Second, economic shocks are likely to have feedback effects. A supply shock in one firm could affect the supply of downstream firms, which eventually could flow back to the original firm. Thus, economic shocks are unlikely to be restricted to follow paths or trails, in which nodes and links are not repeated. This rules out degree centrality. Given that the paths of propagation of economics shocks are unrestricted, and that the Markov transition matrix T serves as a good proxy for contagion probabilities, the long run distribution f_ν of the Markov chain is a suitable measure of contagion risk. It is the stationary distribution that would arise as a shock transitioned from one sector to another an infinite number of times, with no restriction on propagation paths. It measures the likelihood that a firm will receive a random shock that transmits across the network.

[5] [25] provides an overview of network flow measures and their classification based on assumptions about how traffic flows in a network.

[6] See [25] for further discussion on this point

4.3. Corresponding Network Stochastic Discount Factor

In addition to arguing for the validity of our network pricing factor on intuitive grounds, we define a corresponding network stochastic discount factor and show that it has an affine relationship with the network pricing factor, as expected when one relates a stochastic discount factor to a beta pricing model. We formulate a portfolio choice problem where an investor takes into account the mutual exposures between different firms.[7] The agent in our model can be, for example, an institutional investor who considers an firm/industry's exposure to other firms/industries as part of her *ex ante* credit risk assessment at origination.

Our reasoning extends arbitrage pricing theory to incorporate network structure of mutual exposures. As in the classical case, first one establishes that no arbitrage holds by showing the existence of a network stochastic discount factor (SDF). With a network SDF in hand, a cross-sectional regression with returns being the dependent variable becomes a beta pricing model with the regressors being risk premia. We first show the existence of a network SDF and state its comparative static properties with respect to the network structure.

4.3.1. Arbitrage Pricing

To make the analogy clear, we summarize relevant facts of arbitrage pricing theory before proposing a network counterpart. Consider a market with two periods $t = 0, 1$, finitely many securities, and finitely many states. The market is said to admit a *beta pricing model* with factor f if there exists constants R_z and λ such that, for all attainable state-contingent returns R,[8]

$$E_t[R] = R_z + \lambda \frac{Cov(f, R)}{Var(f)}.$$

[7] In practice, mutual exposures is available as part of information disclosed to regulators as firms demonstrate they satisfy capital requirements and to shareholders. In the United States, this is required of firms by Regulation SFAS No. 131 since 1997

[8] Throughout the paper, $E_t[\cdot]$ denotes the conditional expectation operator with respect to information set at time t, $Var(\cdot, \cdot)$ the variance operator, and $Cov(\cdot, \cdot)$ the covariance operator.

If f is a return and there is a risk-free return R_f, then $R_z = R_f$.[9] $\frac{Cov(f,R)}{Var(f)}$ is the risk, or beta, with respect to f, and λ is the risk premium of f-risk. The CAPM is a beta pricing model with f being the market return.

Arbitrage pricing theory is the observation that an approximate beta pricing model can be obtained if the market admits no arbitrage, i.e. there exists a stochastic discount factor (SDF). A SDF is a random variable M such that, for all attainable state-contingent claims R,

$$E[MR] = 1.$$

We assume there is a risk-free return R_f, so that $E[M] = \frac{1}{R_f}$. If a SDF exists, then a cross-sectional regression

$$R_i = E[R_i] + \frac{Cov(f,R)}{Var(f)}(f - E[f]) + \epsilon_i$$

with exogenous and cross-sectionally uncorrelated error term ϵ_i gives an approximate beta pricing model:

$$E[R_i] - R_f = -R_f Cov(f,M)\frac{Cov(f,R_i)}{Var(f)} - \frac{E[M\epsilon_i]}{E[M]}.$$

If the error terms ϵ_i's have uniformly bounded variance, the pricing errors $\delta_i = -\frac{E[M\epsilon_i]}{E[M]}$ tend to zero as the number of securities becomes large.[10] With the idiosyncratic risk ϵ_i diversifiable to zero,

,

$$E[R_i] - R_f \approx -R_f Cov(f,M)\frac{Cov(f,R_i)}{Var(f)}.$$

is a beta pricing model where $-R_f Cov(f,M)$ is the risk premium for the risk factor f and $\frac{Cov(f,R_i)}{Var(f)}$ is the beta.

When an SDF is not formally available, in practice one establishes a factor model by empirically verifying the converse of APT (see e.g. [51] and [31]). In this paper we obtain

[9] More generally, R_z is the expected return of a zero-beta portfolio.
[10] More precisely, $\sum_i \delta_i^2 < \infty$.

a network SDF that only depend on the network structure and corresponds to the pricing factor f_ν. More formally, we decompose the states ω of the world into (ν, ω') where ν captures the network structure and ω' all other aspects of uncertainty in the economy. The network SDF and network pricing factor f_ν are then random variables that depend only on ν with ω' integrated out. The network SDF and network systematic factor are directly related in that they stem from the centrality measure of a firm in the network.

4.3.2. Network SDF

Consider an agent who faces a portfolio problem of allocating his total wealth w among n-assets, indexed by $i = 1, \cdots, n$. Let agent have utility function $u(w_1, \cdots, w_n)$ for w_i invested in asset i. Agent's utility function u is fully general, up to standard regularity assumptions such as quasi-concavity. Therefore u can reflect any heterogeneity of firms, according to the agent's preference. For example, investment in different assets may be pairwise complementary or substitutes.

Suppose the assets are securities issued by n firms with cross-holdings described by network adjacency matrix

$$A = \begin{bmatrix} 0 & a_{12} & \cdots & a_{1n} \\ a_{21} & 0 & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}.$$

The diagonal elements of A are zero. The network structure can reflect a variety of empirical creditor-obligor relationships. For example, the firms can be countries and the network represents cross-holdings of sovereign debts, with the agent choosing a portfolio of sovereign bonds. Another example is a network that represents supplier-consumer relationships. In deciding whether to purchase a security, the investor assesses the firms' consumer and supplier relationships in his risk-return analysis.

Recall our convention that the entry $[a_{ij}]$ correspond to firm i's exposure to firm j.[11] Consider, for example, firm 1, whose exposure profile is $(0, a_{12}, \cdots, a_{1n})$. We assume that for every \$1 the investor allocates to firm 1 a certain amount gets passed to firms $2, \cdots, n$ as determined by mutual exposures. Several empirical scenarios fit this description. For example, firm 1 may need to appropriate part of its equity as risk capital buffering against credit risk. We assume the relative proportion of amounts that get passed by firm 1 to firms $2, \cdots, n$ is determined by firm 1's manager, therefore exogenous to the investor. We also model the relative proportion in a reduced form way and assume it is proportional to the exposures $(a_{12}, \cdots, a_{1n})$.

The agent's maximization problem with respect to firm 1 is therefore

$$\max u(w_1, w_2, \cdots, w_n)$$

subject to the constraint

$$P_h w_1 + P_d(w_2 + \cdots + w_n) = w,\ (w_2, \cdots, w_n) = t(a_{12}, \cdots, a_{1n}),\ t > 0.$$

The prices P_h and P_d are the equilibrium prices paid by the investor per \$1 allocated to firm 1 and its debtors respectively. The two prices might differ because, for example, the investor might demand different returns due to different level of credit risk from amounts lent that gets passed on to other firms. We will see that the relative price $\frac{P_d}{P_h}$ is in fact endogenously determined via the network structure.[12]

Assuming an interior solution exists, in particular, at the optimum $t > 0$. Agent's first order condition with respect to firm 1 is

$$u_2 a_{12} + \cdots + u_n a_{1n} = \frac{P_d}{P_h} u_1,$$

where u_j denotes partial derivative of u with respect to j-th argument, at the investor's optimal portfolio choice with respect to firm 1.

[11] The opposite interpretation of firm i's asset held by firm j is equally available, by replacing the adjacency matrix A with its transpose A'.

[12] The relative price $\frac{P_d}{P_h}$ arises as the Perron-Frobenius root of the network adjacency matrix A.

Therefore, putting the FOC's of investor with respect to all firms together, we have the matrix equation

$$\begin{bmatrix} 0 & a_{12} & \cdots & a_{1n} \\ a_{21} & 0 & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} u_1 \\ \vdots \\ \vdots \\ u_n \end{bmatrix} = \frac{P_d}{P_h} \begin{bmatrix} u_1 \\ \vdots \\ \vdots \\ u_n \end{bmatrix}.$$

If the vector of marginal utilities

$$v = \begin{bmatrix} u_1 \\ \vdots \\ \vdots \\ u_n \end{bmatrix} \tag{4.1}$$

is entry-wise positive, then it must be the Perron-Frobenius eigenvector of the adjacency matrix A corresponding to Perron-Frobenius root $\frac{P_d}{P_h}$.[13]

The centrality vector of marginal utilities v is the network stochastic discount factor. In an asset pricing setting without network considerations, a SDF is agent's state-dependent marginal rate of substitution. Similarly, v is the agent's firm-dependent marginal rate of substitution. A SDF in a dynamic asset pricing model gives the equilibrium relationship between consumption and asset price arising from utility-maximizing agent's behavior and allows assets to be priced by taking consumption as exogenous. In our setting, the additional ingredient of a network structure makes the relative price $\frac{P_d}{P_h}$ endogenous. An SDF is determined by the market structure and independent of agent's attitude towards risk. The same holds for the network SDF v, which is uniquely determined by the network structure and independent of investor preference.

Further analogy with asset pricing theory can be drawn. A SDF is the Radon-Nikodym derivative of an equivalent risk neutral measure after suitable normalization. In our setting, a suitably normalized network SDF v is a *network martingale measure* on firms.[14] Under

[13] The is part of the Perro-Frobenius theorem characterizing the Perron-Frobenuis root. An eigenvalue with an eigenvector whose entries are all strictly positive must be the Perron-Frobenius root.

[14] We assign uniform distribution on firms in the absence of network structure. For strongly connected networks, the probability measure on firms obtained by normalizing v is then absolutely continuous with respect to the uniform distribution with the network SDF v being the Radon-Nikodym derivative.

the risk neutral measure, discounted prices form a martingale and risk-adjusted discounted excess return is therefore zero. In the network setting, rewriting the equation defining v, Equation 4.1, as

$$\begin{bmatrix} 0-1 & \frac{P_h}{P_d}a_{12} & \cdots & \frac{P_h}{P_d}a_{1n} \\ \vdots & & & \\ \frac{P_h}{P_d}a_{n1}-0 & \frac{P_h}{P_d}a_{n2}-0 & \cdots & 0-1 \end{bmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = 0.$$

shows that the expected discounted "excess network return" is equal to zero under the network risk neutral measure.

4.3.3. Affine Relationship with Network Pricing Factor

As outlined in Section 4.3.1, given a SDF M, any affine transformation of M is a pricing factor in a beta pricing model: if $M = a + bf$ for some constants a and b, then for any return R

$$E[R] = \frac{1}{E[M]} - b\frac{1}{E[M]}Cov(f, R)$$

is a beta pricing model with factor f and factor premium $-b\frac{Var(f)}{E[M]}$. Conversely, given a beta pricing model with respect to factor f and expected zero beta return $R_z \neq 0$, there is a corresponding SDF M that is an affine function of f:

$$M = \frac{1}{R_z}(1 - \frac{\lambda}{Var(f)}(f - E[f]))$$

is a SDF. Numerical simulations suggest that a parallel affine relationship also exists between the network SDF v and network pricing factor f_μ, up to small errors. Figure **??** shows the plots of f_μ against v for 1000 financial networks with 60 firms whose mutual exposures are randomly drawn using different classes of distributions.

Diversifiability and network externality The positive relationship between network SDF v and the pricing factor f_ν shown in Figure **??** is in contrast to more familiar beta pricing models, where the factors are themselves returns. This is due to the fact that ex-

posure to the network pricing factor has two countervailing effects. A firm in the network trades off between contagion risk and diversification within the network. Without considering the network structure, diversification is achieved by an idealized portfolio as the number of assets becomes large. In our setting, diversification is relative to the network structure. A firm that is well-diversified by having many connection in the network may also have high network contagion. Depending on the network structure, a firm may more likely to suffer contagion precisely because it is well-connected. The positive relationship between the network structure (therefore T) and the network pricing factor f_ν is a reflects this observation. In a more formal representation where a general state $\omega = (\nu, \omega')$ of the world where ν captures the network structure, a general cross-sectional shock of the economy may be uncorrelated at the level of ω but not necessarily uncorrelated once ω' is integrated out and the network factor is isolated.

4.3.4. Comparative Statics of Network SDF

As further analysis on properties of network SDF v, we examine its comparative statics with respect to network structure. The comparative statics properties reinforces results from Section 4.3. In addition, using the positive affine relationship with f_ν, they shed light on on diversification vs. susceptibility to adverse shock trade-off captured by the network pricing factor.

A network is said to be *strongly connected* if any two firms can be connected by a directed path.

Proposition 4.3.1. *In a strongly connected financial network, suppose firm i increases its exposures in other firms. Let v and v' be the network SDF before and after the increase in firm i's exposures profile, then*

$$\log v_i' - \log v_i > \log v_j' - \log v_j, \quad \forall j \neq i.$$

Proposition 4.3.1 gives a partial monotonicity characterization of the highly nonlinear relationship between network SDF and network structure. In our setting, the investor

demands higher marginal utility as it becomes more central—e.g. when the success of a project depends on the performance of multiple subsidiaries, etc. The same is applicable to network pricing factor.

Next we have a smoothness result, i.e. there are no abrupt changes in network SDF as the network structure varies. Since any financial network can be arbitrarily approximated by strongly connected networks[15], the smoothness property in turn implies assuming strong connectedness is without loss of generality.

Proposition 4.3.2. *Let Δr be a vector with non-negative entries. Consider the perturbations of the network where are firm i increase his exposures by $\eta \cdot \Delta r$ with network SDF $v(\eta)$. Then there exists $\delta > 0$ such that, on $[0, \delta)$, $v_j(\eta)$ is a decreasing differentiable function of η, for $j \neq i$.*

Example 4.3.3. *We give an numerical example of the smoothness property stated in Proposition 4.3.2. Consider a financial network with 4 firms and a ring configuration, depicted in Figure 4.2(a). The network adjacency matrix is*

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

This give network SDF

$$v = \begin{bmatrix} \frac{1}{4} \\ \frac{1}{4} \\ \frac{1}{4} \\ \frac{1}{4} \end{bmatrix}$$

Suppose now firm 1 increases its exposure to firm 2 by \$$\epsilon$, as shown in Figure 4.2. The resulting network SDF from this perturbation of the network structure is

[15] More formally, we identify a network with its adjacency matrix. Under this identification, strongly connected networks are dense in the class of networks under consideration, in the unique locally convex topology on matrices.

$$\begin{bmatrix} \lambda^3 \\ 1 \\ \lambda \\ \lambda^2 \end{bmatrix}$$

where $\lambda = (1+\epsilon)^{\frac{1}{4}} > 1$. *As the one who assumed additional exposure, firm 1's network SDF gets the most increase by a factor of* λ^3. *The effects on the other firms is determined by the degree of connectedness to firm 1. The second highest increase in network SDF is incurred by firm 1's immediate debtor, firm 4, with successively less increases for firms 3 and 2.*

We have a natural extension of Proposition 4.3.2 to the case where more than one firm increases their exposures.

Proposition 4.3.4. *In a strongly connected network, if a group of firms* $i_1, \cdots, i_m$ *increase their exposures, then*

$$\forall j \notin \{i_1, \cdots, i_m\},\ \log v'_j - \log v_j < \max_{i \in \{i_1, \cdots, i_m\}} \log v'_i - \log v_i.$$

Finally, we formally consider the case where one firms increases its exposures in the network while a second one undergoes a reduction in exposure:

Proposition 4.3.5. *Consider a strongly connected financial network where firm* i *firm* j *increases and decreases hers. Then for all firms* k,

(i) If $\frac{P_h}{P_d} \leq \frac{P'_h}{P'_d}$, $\log v'_k - \log v_k \leq \log v'_i - \log v_i$.

(ii) If $\frac{P_h}{P_d} \geq \frac{P'_h}{P'_d}$, $\log v'_k - \log v_k \geq \log v'_j - \log v_j$.

(iii) If $\frac{P_h}{P_d} = \frac{P'_h}{P'_d}$, $\log v'_j - \log v_j \leq \log v'_k - \log v_k \leq \log v'_j - \log v_j$.

Example 4.3.6. *This numerical example illustrates Proposition 4.3.5. Suppose now firm 1 incurs* $\$\epsilon$ *more exposure while firm 3 reduces its exposure by* η, *as shown in Figure* **??**.

This two changes have competing effects on the network SDF. Firm 1's additional exposure causes his network SDF to increase, with attendant propagation effects downstream to firms for whom he is a obligor. Firm 3's reduction has the opposite effect. If ϵ is large relative to η, i.e. there is a net increase in exposures in the network, then the relative prices $\frac{P_h}{P_d}$ of deposits increases. Deposits become less attractive across the network as net exposure increases. This is case (i) of Proposition 2. Conversely, if ϵ is small relative to η, this results in case (ii). The adjacency matrix is

$$A = \begin{bmatrix} 0 & 1+\epsilon & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1-\eta \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

$\frac{P_h}{P_d} = (1+\epsilon)^{\frac{1}{4}}(1-\eta)^{\frac{1}{4}}$. *The network SDF is, up to normalization,*

$$\begin{bmatrix} (1+\epsilon)^{\frac{1}{4}}(1-\eta)^{\frac{1}{4}} \\ \frac{(1-\eta)^{\frac{1}{2}}}{(1+\epsilon)^{\frac{1}{2}}} \\ \frac{(1-\eta)^{\frac{3}{4}}}{(1+\epsilon)^{\frac{1}{4}}} \\ 1 \end{bmatrix}.$$

If $(1+\epsilon)(1-\eta) > 1$, then firm 1's network SDF undergoes the most relative increase. Although Theorem 4.8 In this particular case firm 3's network SDF decreases, in relative terms. Conversely, if $(1+\epsilon)(1-\eta) < 1$, the effect is in the opposite direction.

4.4. Summary

We have provided a microfouned network pricing factor that captures the trade-off between diversification benefits and exposure to contagion risk. A priori diversifiable risk need not remain diversifiable once non-network uncertainty is integrated out and only the network structure remains. A well-connected firm with many connections in the network

can very well have high network risk. In the point argued formally in Section 4.3.3, however, we have shown that diversifiability does compensate for exposure to contagion risk.

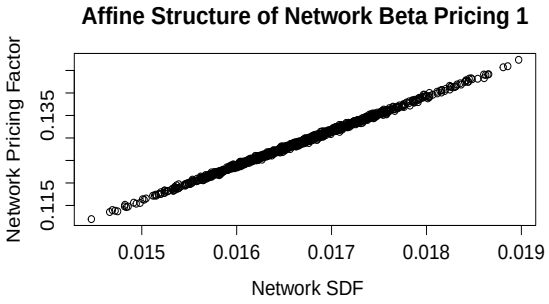
Our model can tested empirically, which is planned for future research, using the classical Fama-French-MacBeth approach—first sorting firms and form portfolios based on network pricing factor then performing cross-sectional regressions along the lines of [51]—using, for example, a consumer-supplier network.[16] Some empirical flexibility is embedded in our formulaton. For example, depending on convention used in writing down the adjacency matrix, the network risk factor can reflect the vulnerability of a firm to contagion risk either as a supplier or consumer.

Another empirical test ground for the model is credit spreads of firms in a financial network of cross holding. Conventional credit risk approach posits that an industry that replies on just a few suppliers has greater credit risk than an industry with multiple suppliers. The reason given is that, from the supply side, monopoly suppliers have higher bargaining power to capture a higher portion of total surplus—sometimes dubbed *the power of supplier* by credit analysts. A larger number of (more competitive) suppliers helps an industry keep input prices in check. Similarly, on the demand size, an industry that relies on a few consumers is subject to more powerful forces—e.g. the consumers demanding better product and lower price—driving it toward s zero profit. Such linear portfolio-level analysis ignores macro-level nonlinear network feedback effects captured by the network risk factor.

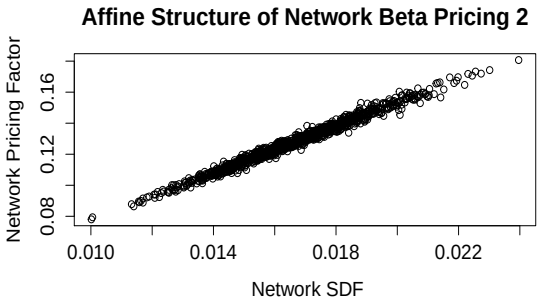
[16] US Regulation SFAS No. 131 requires firms to report selected information about operating segments in interim financial reports issued to shareholders. In particular, firms are required to disclose certain financial information for any industry segment that comprised more than 10% of consolidated yearly sales, assets, or profits, and the identity of any customer representing more than 10% of the total reported sales.

Figure 4.1: Affine Structure

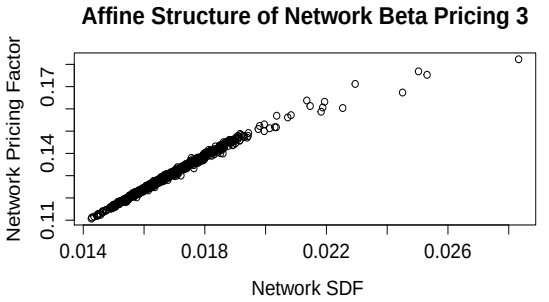
(a)χ^2 distribution



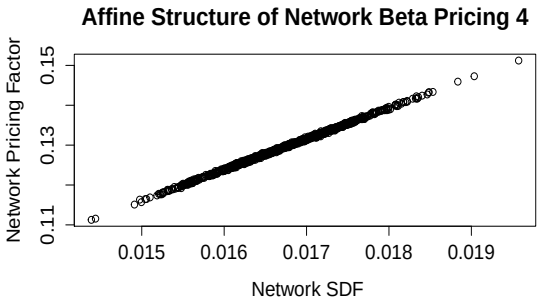
(b)Exponential distribution

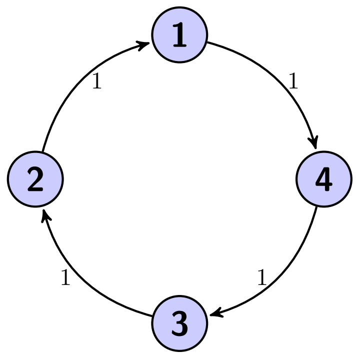


(c)Power law distribution

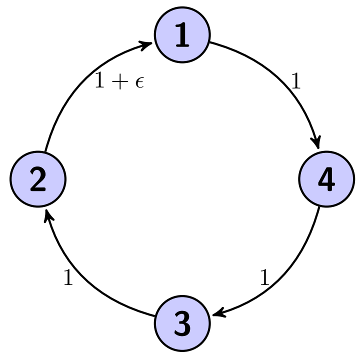


(d)Gamma distribution





(a)Before firm 1 increases its exposure to firm 2



(b)After firm 1 increases its exposure to firm 2

Figure 4.2: A Ring Financial Network.

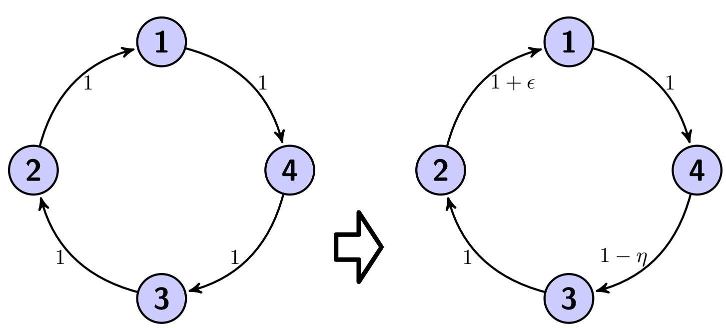


Figure 4.3: A Simultaneous Adjustment of Two Firms' Exposures in the Ring Network Configuration.

Chapter 5

Conclusion

This thesis discusses three interweaved topics of modern financial economics—high frequency econometrics, market microstructure, and financial networks. The new research frontiers in both market microstructure and financial networks have been brought about by fundamental, and ongoing, technological innovation that drives an increasing interconnected global financial market. In addition to shifting market behavior and economic perspectives, the same technological innovation also makes available high frequency data that demands new econometrics methodology. It is an exciting time to be a student of financial economics. We hope that each of the three essays has made a contribution to their respective topics and will serve as stepping stones for future research.

Bibliography

[1] Daron Acemoglu, Vasco M Carvalho, Asuman Ozdaglar, and Alireza Tahbaz-Salehi. The network origins of aggregate fluctuations. *Econometrica*, 80(5):1977–2016, 2012.

[2] Kenneth R Ahern. Network centrality and the cross section of stock returns. *Available at SSRN 2197370*, 2013.

[3] Kenneth R Ahern and Jarrad Harford. The importance of industry links in merger waves. *Journal of Finance*, 69(2):527–576, 2014.

[4] Yacine Ait-Sahalia and Jean Jacod. Estimating the degree of activity of jumps in high frequency data. *Annals of Statistics*, 37:2202–2244, 2009.

[5] Yakov Amihud and Haim Mendelson. Dealership market: Market making with inventory. *Journal of Financial Economics*, 8:31–53, 1980.

[6] Torben G. Andersen, Tim Bollerslev, and Francis X. Diebold. Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *Review of Economics and Statistics*, 89:701–720, 2007.

[7] Torben G. Andersen, Tim Bollerslev, and Dobrislav Dobrev. No-arbitrage semimartingale restrictions for continuous-time volatility models subject to leverage effects, jumps and i.i.d. noise: Theory and testable distributional implications. *Journal of Econometrics*, 138:125–180, 2007.

[8] Torben G. Andersen, Tim Bollerslev, Per Frederiksen, and Morten Ørregaard Nielsen. Continuous-time models, realized volatilities, and testable distributional implications for daily stock returns. *Journal of Applied Econometrics*, 25:233–261, 2010.

[9] Torben.G. Andersen, Tim Bollerslev, Diebold Francis.X., and Vega Clara. Micro effects of macro announcements: Real-time price discovery in foreign exchange. *American Economic Review*, 93:38–62, 2003.

[10] Donald WK Andrews. Tests for parameter instability and structural change with unknown change point. *Econometrica*, 61(4):821–856, 1993.

[11] James Angel. Tick size, share prices, and stock splits. *Journal of Finance*, 52:655–681, 1997.

[12] Nimalan Arinaminpathy, Sujit Kapadia, and Robert M. May. Size and complexity in model financial systems. *Proceedings of the National Academy of Sciences*, 109(45):18338–18343, 2012.

[13] Jushan Bai and Pierre Perron. Estimating and testing linear models with multiple structural changes. *Econometrica*, 66(1):47–78, 1998.

[14] Clifford A. Ball, Walter N. Torous, and Adrian E. Tschoegl. The degree of price resolution: The case of the gold market. *Journal of Futures Markets*, 5:29–43, 1985.

[15] Ole E. Barndorff-Nielsen and Neil Shephard. Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics*, 2:1–37, 2004.

[16] Ole E. Barndorff-Nielsen and Neil Shephard. Econometrics of testing for jumps in financial economics using bipower variation. *Journal of Financial Econometrics*, 4:1–30, 2006.

[17] Ole E. Barndorff-Nielsen and Neil Shephard. Impact of jumps on returns and realised variances: Econometric analysis of time-deformed Levy processes. *Journal of Econometrics*, 131:217–252, 2006.

[18] Abraham Berman and Robert J Plemmons. Nonnegative matrices. *The Mathematical Sciences, Classics in Applied Mathematics,*, 9, 1979.

[19] Oleg Vladimirovich Besov, Valentin Petrovich IlÊźin, and Sergeĭ Mikhaĭlovich NikolÊźskiĭ. *Integral Representations of Functions and Embedding Theorems*. Halsted Press, 1978.

[20] Bruno Biais and Paul Woolley. High frequency trading. *Manuscript, Toulouse University, IDEI*, 2011.

[21] Herman J Bierens. Kernel estimators of regression functions. In *Advances in Econometrics: Fifth World Congress, Volume 1*, 1987.

[22] Monica Billio, Mila Getmansky, Andrew W Lo, and Loriana Pelizzon. Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of Financial Economics*, 104(3):535–559, 2012.

[23] Geir H. Bjønnes and Dagfinn Rime. Dealer behavior and trading systems in foreign exchange markets. *Journal of Financial Economics*, 75:571–605, 2005.

[24] James N Bodurtha and Nelson C Mark. Testing the CAPM with time-varying risks and returns. *Journal of Finance*, 46(4):1485–1505, 1991.

[25] Stephen P Borgatti and Martin G Everett. Models of core/periphery structures. *Social Networks*, 21(4):375 – 395, 2000.

[26] Jonathan Brogaard, Terrence Hendershott, and Ryan Riordan. High frequency trading and price discovery. *Technical Report, University of California Berkeley*, 2012.

[27] Jane P Brown, Haiyan Song, and Alan McGillivray. Forecasting UK house prices: A time varying coefficient approach. *Economic Modelling*, 14(4):529–548, 1997.

[28] Lawrence D Brown, Mark G Low, et al. Asymptotic equivalence of nonparametric regression and white noise. *Annals of Statistics*, 24(6):2384–2398, 1996.

[29] Zongwu Cai. Trending time-varying coefficient time series models with serially correlated errors. *Journal of Econometrics*, 136(1):163–188, 2007.

[30] Zongwu Cai, Qi Li, and Joon Y Park. Functional-coefficient models for nonstationary time series data. *Journal of Econometrics*, 148(2):101–113, 2009.

[31] Mark M Carhart. On persistence in mutual fund performance. *Journal of finance*, 52(1):57–82, 1997.

[32] Alain Chaboud, Benjamin Chiquoine, Erik Hjalmarsson, and Clara Vega. Rise of the machines: Algorithmic trading in the foreign exchange market. *Journal of Finance*, 69:2045–2084, 2014.

[33] Alain P. Chaboud, Sergey V. Chernenko, Edward Howorka, Raj S. Krishnasami Iyer, David Liu, and Jonathan H Wright. The high-frequency effects of U.S. macroeconomic data releases on prices and trading activity in the global interdealer foreign exchange market. *Board of Governors of the Federal Reserve System, International Finance Discussion Papers*, Number 823, 2004.

[34] Bin Chen and Yongmiao Hong. Testing for smooth structural changes in time series models via nonparametric regression. *Econometrica*, 80(3):1157–1183, 2012.

[35] Son-Nan Chen and Arthur J Keown. Risk decomposition and portfolio diversification when beta is nonstationary: a note. *Journal of Finance*, 36(4):941–947, 1981.

[36] Kim Christensen, Roel C. A. Oomen, and Mark Podolskij. Fact or friction: Jumps at ultra high frequency. *Journal of Financial Economics*, 114:576–599, 2014.

[37] William G. Christie and Paul H. Schults. Why do NASDAQ market makers avoid odd-eighth quotes? *Journal of Finance*, 49:1813–1840, 1994.

[38] Z Ciesielski. Hölder conditions for realizations of Gaussian processes. *Transactions of the American Mathematical Society*, 99(3):403–413, 1961.

[39] Albert Cohen, Ingrid Daubechies, and Pierre Vial. Wavelets on the interval and fast wavelet transforms. *Applied and Computational Harmonic Analysis*, 1(1):54–81, 1993.

[40] Pierre Collin-Dufresne and Vyacheslav Fos. Do prices reveal the presence of informed trading? *Journal of Finance*, 70(4):1555–1582, 2015.

[41] Thomas E. Copeland and Dan Galai. Information effects on the bid-ask spread. *Journal of Finance*, 38:1457–1469, 1983.

[42] Ingrid Daubechies. Orthonormal bases of compactly supported wavelets. *Communications on Pure and Applied Mathematics*, 41(7):909–996, 1988.

[43] Robert M De Jong and James Davidson. The functional central limit theorem and weak convergence to stochastic integrals I. *Econometric Theory*, 16(05):621–642, 2000.

[44] David L Donoho. De-noising by soft-thresholding. *IEEE Transactions on Information Theory*, 41(3):613–627, 1995.

[45] David L Donoho. Nonlinear solution of linear inverse problems by wavelet–vaguelette decomposition. *Applied and Computational Harmonic Analysis*, 2(2):101–126, 1995.

[46] David L Donoho, Iain M Johnstone, et al. Minimax estimation via wavelet shrinkage. *The Annals of Statistics*, 26(3):879–921, 1998.

[47] David Easley, Marcos Lopez de Prado, and Maureen O'Hara. Discerning information from trade data. *NBER Working Paper, http://ssrn.com/abstract=1989555*, 2015.

[48] David Easley and Maureen O'Hara. Price, trade size, and information in securities markets. *Journal of Financial Economics*, 19:69–90, 1987.

[49] Martin DD Evans. Expected returns, time-varying risk, and risk premia. *Journal of Finance*, 49(2):655–679, 1994.

[50] Eugene F. Fama and Kenneth R. French. Permanent and temporary components of stock prices. *Journal of Political Economy*, 96:246–273, 1988.

[51] Eugene F Fama and Kenneth R French. Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, 33(1):3–56, 1993.

[52] Eugene F Fama and Kenneth R French. A five-factor asset pricing model. *Journal of Financial Economics*, 116(1):1–22, 2015.

[53] Yanqin Fan and Ramazan Gençay. Unit root tests with wavelets. *Econometric Theory*, 26(05):1305–1331, 2010.

[54] Jon Faust, John H. Rogers, Shing-Yi B. Wang, and Jonathan H. Wright. The high-frequency response of exchange rates and interest rates to macroeconomic announcements. *Journal of Monetary Economics*, 54:1051–1068, 2007.

[55] Céline Gauthier, Alfred Lehar, and Moez Souissi. Macroprudential capital requirements and systemic risk. *Journal of Financial Intermediation*, 21(4):594–618, 2012.

[56] IzrailÊź M GelÊźfand and N Yao Vilenkin. *Generalized functions*. Academic press, 1964.

[57] Ramazan Gençay, Miche Dacorogna, Ulrich A. Müller, Olivier Pictet, and Richard Olsen. *An Introduction to High-Frequency Finance*. Academic Press, 2001.

[58] Ramazan Gençay and Daniele Signori. Multi-scale tests for serial correlation. *Journal of Econometrics*, 184(1):62–80, 2015.

[59] Eric Ghysels. On stable factor structures in the pricing of risk: Do time-varying betas help or hurt? *Journal of Finance*, 53(2):549–573, 1998.

[60] Scott Gibson, Rajdeep Singh, and Vijay Yerramilli. The effect of decimalization on the components of the bid-ask spread. *Journal of Financial Intermediation*, 12:121–148, 2003.

[61] IV Girsanov. On transforming a certain class of stochastic processes by absolutely continuous substitution of measures. *Theory of Probability & Its Applications*, 5(3):285–301, 1960.

[62] Lawrence R. Glosten and Paul R. Milgrom. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics*, 14:71–100, 1985.

[63] Michael A. Goldstein and Kenneth A. Kavajecz. Eighths, sixteenths, and market depth: Changes in tick size and liquidity provision on the NYSE. *Journal of Financial Economics*, 56:125–149, 2000.

[64] C. Goodhart and R. Curcio. Asset price discovery and price clustering in the foreign exchange market. *Unpublished paper, London School of Economics*, 1990.

[65] Hany S Guirguis, Christos I Giannikos, and Randy I Anderson. The US housing market: Asset pricing forecasts using time varying coefficients. *The Journal of Real Estate Finance and Economics*, 30(1):33–53, 2005.

[66] Owain Gwilym, Andrew Clare, and Stephen Thomas. Extreme price clustering in the London equity index futures and options markets. *Journal of Banking & Finance*, 22:1193–1206, 1998.

[67] Lawrence E. Harris. Minimum price variations, discrete bid-ask spreads, and quotation sizes. *Review of Financial Studies*, 7:149–178, 1994.

[68] Joel Hasbrouck. The summary informativeness of stock trades: An econometric analysis. *Review of Financial Srudies*, 4:571–595, 1991.

[69] Joel Hasbrouck and Gideon Saar. Low-latency trading. *Journal of Financial Markets*, 16:646–679, 2013.

[70] Trevor J Hastie and Robert J Tibshirani. *Generalized Additive Models*, volume 43. CRC Press, 1990.

[71] Terrence Hendershott, Charles M. Jones, and Albert J. Menkveld. Does algorithmic trading improve liquidity? *Journal of Finance*, 66:1–33, 2011.

[72] Volker Herren. Lévy-type processes and Besov spaces. *Potential Analysis*, 7(3):689–704, 1997.

[73] Thomas Ho and Hans Stoll. Optimal dealer pricing under transactions and return uncertainty. *Journal of Financial Economics*, 9:47–73, 1981.

[74] Daniel A. Hojman and Adam Szeidl. Core and periphery in networks. *Journal of Economic Theory*, 139(1):295 – 309, 2008.

[75] Yongmiao Hong and Chihwa Kao. Wavelet-based testing for serial correlation of unknown form in panel models. *Econometrica*, 72(5):1519–1563, 2004.

[76] Roger D. Huang and Hans R. Stoll. The components of the bid-ask spread: A general approach. *Review of Financial Studies*, 10:995–1034, 1997.

[77] Roger D. Huang and Hans R. Stoll. Tick size, bid-ask spreads, and market structure. *The Journal of Financial and Quantitative Analysis*, 36:503–522, 2001.

[78] Xin Huang and George Tauchen. The relative contribution of jumps to total price variance. *Journal of Financial Econometrics*, 3:456–499, 2005.

[79] Kiyosi Itô, Makiko Nisio, et al. On the convergence of sums of independent Banach space valued random variables. *Osaka Journal of Mathematics*, 5(1):35–48, 1968.

[80] Jean Jacod, Yingying Li, Per A. Mykland, Mark Podolskij, and Mathias Vetter. Microstructure noise in the continuous case: The pre-averaging approach. *Stochastic Processes and their Applications*, 119:2249–2276, 2009.

[81] Yitzhak Katznelson. *An Introduction to Harmonic Analysis*. Cambridge University Press, 2004.

[82] Dennis Kristensen. Non-parametric detection and estimation of structural change. *The Econometrics Journal*, 15(3):420–461, 2012.

[83] Albert S. Kyle. Continuous auctions and insider trading. *Econometrica*, 53:1315–1335, 1985.

[84] Gabriel Lang and Philippe Soulier. Convergence de mesures spectrales aléatoires et applications à des principes d'invariance. *Statistical Inference for Stochastic Processes*, 3(1-2):41–51, 2000.

[85] Harris Lawrence. Stock price clustering and discreteness. *Review of Financial Studies*, 4:389–415, 1991.

[86] Jin Lee and Yongmiao Hong. Testing for serial correlation of unknown form using wavelet methods. *Econometric Theory*, 17(02):386–423, 2001.

[87] Paul Lévy. Le mouvement brownien. *Mémorial des sciences mathématiques*, 126:1–84, 1954.

[88] Robert E Lucas Jr. Econometric policy evaluation: A critique. In *Carnegie-Rochester Conference Series on Public Policy*. Elsevier, 1976.

[89] Richard Lyons. Tests of microstructural hypotheses in the foreign exchange market. *Journal of Financial Economics*, 39:321–351, 1995.

[90] Yves Meyer. *Wavelets and Operators*. Cambridge University Press, 1995.

[91] Joseph G Nellis and J Andrew Longbottom. An empirical analysis of the determination of house prices in the United Kingdom. *Urban Studies*, 18(1):9–21, 1981.

[92] Maureen O'Hara. High frequency market microstructure. *Journal of Financial Economics*, 116(2):257–270, 2015.

[93] Wataru Ohta. An analysis of intraday patterns in price clustering on the Tokyo Stock Exchange. *Journal of Banking & Finance*, 30:1023–1039, 2006.

[94] Richard Payne. Informed trade in spot foreign exchange markets: An empirical investigation. *Journal of International Economics*, 61:307–329, 2003.

[95] Pierre Perron and Yohei Yamamoto. A note on estimating and testing for multiple structural changes in models with endogenous regressors via 2SLS. *Econometric Theory*, 30(02):491–507, 2014.

[96] MS Pinsker. Optimal filtration of square-integrable signals in Gaussian noise. *Prob. Info. Transmission*, 16(2):120–133, 1980.

[97] Mark Podolskij and Mathias Vetter. Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps. *Bernoulli*, 15:599–924, 2009.

[98] Mark Podolskija and Mathias Vetter. Bipower-type estimation in a noisy diffusion setting. *Stochastic Processes and their Applications*, 119:2803–2831, 2009.

[99] James O Ramsay. *Functional Data Analysis*. Wiley Online Library, 2006.

[100] Peter M Robinson. Root-N-consistent semiparametric regression. *Econometrica*, 56(4):931–954, 1988.

[101] Peter M Robinson. *Nonparametric Estimation of Time-Varying Parameters*. Springer, 1989.

[102] Richard Roll. A simple implicit measure of the effective bid-ask spread in an efficient market. *Journal of Finance*, 39:1127–1139, 1984.

[103] Anatoly B. Schmidt. Ecology of the modern institutional spot FX: The EBS market in 2011. *SSRN: http://ssrn.com/abstract=1984070*, 2012.

[104] Paul Schultz. Stock splits, tick size, and sponsorship. *Journal of Finance*, 55:429–450, 2000.

[105] Hans Stoll. The supply of dealer services in security markets. *Journal of Finance*, 33:1133–1151, 1978.

[106] Murad S Taqqu. Weak convergence to fractional Brownian motion and to the Rosenblatt process. *Probability Theory and Related Fields*, 31(4):287–302, 1975.

[107] Hans Triebel. *Interpolation Theory, Function spaces, Differential Operators*. North Holland, 1978.

[108] Bonnie F. Van Ness, Robert A. Van Ness, and Richard S. Warr. Is the adverse selection component really higher on the NYSE/Amex than on the Nasdaq? *Journal of Business Finance & Accounting*, 29:807–824, 2002.

[109] Alfred Wong and Tom Fong. Analysing interconnectivity among economies. Technical report, Hong Kong Monetary Authority Working Paper, 2010.

[110] Jeffrey M Wooldridge and Halbert White. Some invariance principles and central limit theorems for dependent heterogeneous processes. *Econometric Theory*, 4(02):210–230, 1988.

Appendix A

Proofs for Chapter 2

Proof of Theorem 2.2.1: Let $\kappa(t) = (\beta', \gamma(t))'$. We give an argument for consistent estimation of $\kappa(t_0)$ for any $t_0 \in (0,1)$ at which κ is continuous.

Assume without loss of generality that the K has support $[-1,1]$ so that $K_h(\cdot - t_0)$ has support $[t_0 - h, t_0 + h]$. Then

$$\|\hat{\kappa}(t_0) - \kappa(t_0)\| \leq \underbrace{\|(\frac{1}{n}\sum_{t=1}^{n} K_{h_n}(t-t_0)X_tX_t')^{-1}(\frac{1}{n}\sum_{t=1}^{n} K_{h_n}(t-t_0)X_tX_t'1_{\{|t-t_0|\leq h_n\}}\delta_{h_n}\|}_{R_1}$$
$$+ \underbrace{\|(\frac{1}{n}\sum_{t=1}^{n} K_{h_n}(t-t_0)X_tX_t')^{-1}(\frac{1}{n}\sum_{t=1}^{n} K_{h_n}(t-t_0)X_t\epsilon_t\|}_{R_2},$$

where, by continuity of κ at t_0, $\delta_{h_n} \to 0$ as $h_n \to 0$.

First we have

$$R_1 \leq \|(\frac{1}{n}\sum_{t=1}^{n} K_{h_n}(t-t_0)X_tX_t')^{-1}\|\|\frac{1}{n}\sum_{t=1}^{n} K_{h_n}(t-t_0)X_tX_t'\|\|\delta_{h_n}\|. \tag{A.1}$$

By Assumption 1(ii), $(K_{h_n}(t-t_0)X_tX_t')$ is also α-mixing, with size at least that of (X_t). For a fixed bandwidth h, by boundedness of K_h and Assumption 1(iii), $(K_h(t-t_0)x_{i,t}x_{j,t})$ is uniformly bounded in $L^{2+2\delta}(\mathbb{R})$ for some $\delta > 0$. Therefore by Weak Law of Large Numbers,

$$\frac{1}{n}\sum_{t=1}^{n} K_h(t-t_0)X_tX_t' - \frac{1}{n}\sum_{t=1}^{n} K_h(t-t_0)E[X_tX_t'] \xrightarrow{p} 0.$$

Since $(x_{i,t}x_{j,t})$ is also uniformly bounded in $L^1(\mathbb{R})$ by Jensen's inequality, $\|E[X_tX_t']\| \in O(1)$.[1] It follows from the triangle inequality that the convex sum $\frac{1}{n}\sum_{t=1}^n K_h(t-t_0)E[X_tX_t'] \in O(1)$. This shows that

$$\frac{1}{n}\sum_{t=1}^n K_h(t-t_0)X_tX_t' \in O_p(1).$$

A diagonalization argument then shows that, as $n \to \infty$ and $h_n \to 0$,

$$\frac{1}{n}\sum_{t=1}^n K_{h_n}(t-t_0)X_tX_t' \in O_p(1).$$

Also, by Assumption 1(iv), for all h, $\inf_n \|\frac{1}{n}\sum_{t=1}^n K_h(t-t_0)E[X_tX_t']\| > \eta > 0$. This implies $(\frac{1}{n}\sum_{t=1}^n K_h(t-t_0)E[X_tX_t'])^{-1} \in O(1)$. In turn,

$$(\frac{1}{n}\sum_{t=1}^n K_h(t-t_0)X_tX_t')^{-1} - (\frac{1}{n}\sum_{t=1}^n K_h(t-t_0)E[X_tX_t'])^{-1} \xrightarrow{p} 0$$

implies $\frac{1}{n}\sum_{t=1}^n K_{h_n}(t-t_0)X_tX_t')^{-1} \in O_p(1)$. Therefore $R_1 \in o_p(1)$. Similar argument shows that $R_2 \in o_p(1)$. This proves the theorem.

Lemma A.1 for Theorem 2.4.4 For $B^{\alpha}_{p,q}(L)$ with space of wavelet coefficients $\Theta \subset l^2(\mathbb{N})$, [44] considers a sequence of truncated Gaussian sequence models $\mathcal{M}_n$

$$X_j = \theta_j + e_j, \text{ where } e_j \sim \mathcal{N}(0, \frac{\sigma^2}{n}),\ 1 \le j \le n,$$

where for each n, $(\theta_j) \in \Theta \cap \mathbb{R}^n$ (as it is irrelevant for the argument in this case, we suppress the double indices for wavelet coefficients). $\hat{\gamma}_n$ is constructed by estimating the n coefficients. We show that this is without loss of generality for our global L^2-formulation. As any $\gamma \in L^2[0,1]$ can be approximated by its truncations γ_n, it is without loss of generality to estimate γ_n. Furthermore, estimating γ_n from its first n empirical coefficients does not increase risk: In the filtering model, let $\mathbf{P}_\gamma$ and $\mathbf{P}_0$ be the probability measures on $C[0,1]$ corresponding to the processes $\mathbf{G} = dG_t = \gamma_n dt + \frac{1}{\sqrt{n}}dB_t$ and dB_t respectively. By Girsanov's Theorem ([61]), the likelihood ratio is

$$\frac{d\mathbf{P}_\gamma}{d\mathbf{P}_0} = e^{\frac{1}{n}\sum_{j=1}^n \theta_j X_j - \frac{1}{2n}\sum_{j=1}^n \theta_j^2},$$

which is measurable with respect to the first n empirical coefficients $X_1, \cdots, X_n$. By Jensen's inequality, for any estimator $\hat{\theta}_j(\mathbf{G})$ where $1 \le j \le n$,

[1] Here we use the uniqueness of locally convex topology on a finite dimensional vector space. Boundedness entry-wise is equivalent to boundedness in operator norm.

$$\begin{aligned}\mathbb{E}_\gamma[(\hat{\theta}_j(\mathbf{F})-\theta_j)^2] &= \mathbb{E}_0[\frac{d\mathbf{P}_\gamma}{d\mathbf{P}_0}(\mathbf{F})(\hat{\theta}_j(\mathbf{F})-\theta_j)^2] \\ &= \mathbb{E}_0[\mathbb{E}_0[(\hat{\theta}_j(\mathbf{F})-\theta_j)^2|X_1,\cdots,X_n]\frac{d\mathbf{P}_\gamma}{d\mathbf{P}_0}] \\ &\geq \mathbb{E}_0[(\bar{\theta}_j(X_1,\cdots,X_n)-\theta_j)^2\frac{d\mathbf{P}_\gamma}{d\mathbf{P}_0}] \\ &= \mathbb{E}_\gamma[(\bar{\theta}_j(X_1,\cdots,X_n)-\theta_j)^2],\end{aligned}$$

where $\bar{\theta_j}(X_1,\cdots,X_n)=\mathbb{E}_0[\hat{\theta}_j(\mathbf{F})|X_1,\cdots,X_n]$. Therefore Theorem 1.1 and Theorem 1.2 in [44] imply Theorem 2.4.4.

The Lévy-Ciesielski-Itô construction of Brownian motion This construction was used in passing from the filtering model to the Gaussian sequence model in Section 2.4.1 and generalized to fractional Brownian motion en route to proving Theorem 2.4.5. Let $\epsilon'_j \sim \mathcal{N}(0,1)$ be i.i.d. random variables defined on a probability space $(\Omega,\mathcal{F},P)$. Let $\mathcal{H}$ be the Hilbert subspace of $L^2(\Omega,\mathcal{F},P)$ generated by $\{\epsilon'_j\}$. Define an Itô isometry by

$$\epsilon'_j \in \mathcal{H} \stackrel{\Psi}{\mapsto} \psi_j \in L^2[0,1].$$

The resulting stochastic process $t \mapsto \Psi(1_{[0,t]})$ has the same finite dimensional distributions as standard Brownian motion:

1 The increment $\Psi(1_{[0,t]})-\Psi(1_{[0,s]})=\Psi(1_{[s,t]})$ for any $0 \leq s < t \leq 1$ is distributed $\mathcal{N}(0,t-s)$, being the mean square limit of normal random variables.

2 Two increments $\Psi(1_{[s,t]})$ and $\Psi(1_{[s',t']})$ are uncorrelated, therefore independent by normality.

It then follows from standard facts that $\Psi(1_{[0,t]})$ has a modification with almost surely continuous sample paths.

Definition of fractional Brownian motion We recall here the precise definition of the fractional Brownian motion, which is the limit process for the Functional Central Limit Theorem used in Section 2.4.1. The fractional Brownian motion with Hurst exponent H $B_H(t), t \in \mathbb{R}$, is a zero-mean Gaussian process with covariance function

$$\gamma(s,t)=\frac{V_H}{2}(|s|^{2H}+|t|^{2H}-|t-s|^{2H})$$

where

$$V_H = var(B_H(1)) = \frac{-\Gamma(2-2H)\cos(\pi H)}{\pi H(2H-1)},$$

and $\Gamma(\cdot)$ is the gamma function.

Proof of Theorem 2.4.5: Fix a wavelet basis $\{\psi_{jk}\}$ that is continuously differentiable up to $r > H + \frac{3}{2}$ times. Let $\Delta = \frac{d^2}{dx^2}$ be the Laplace operator on $[0,1]$. For Hurst exponent $H \in (\frac{1}{2}, 1)$, the operator $K_H = (-\Delta)^{H+\frac{1}{2}}$ is the reproducing kernel of the reproducing kernel Hilbert space of dB_t^H.[2] The functions $K_H^{-\frac{1}{2}}\psi_{jk}$ diagonalizes K_H, which gives a Karhunen-Loéve decomposition of fractional Brownian motion

$$B_t^H = \sum_{jk} w_{jk} K_H^{-\frac{1}{2}} \psi_{jk}$$

where $\{w_{jk}\}$ is a Gaussian white noise. Define

$$v_{jk} = \frac{d}{dt} K_H^{-\frac{1}{2}} \psi_{jk} = (-\Delta)^{\frac{1}{4}-\frac{H}{2}} \psi_{jk}.$$

Then we have a representation:

$$dB_t^H = \sum_{jk} w_{jk} v_{jk}(t) dt.$$

The random variables e_{jk} in the statement of the theorem can be expressed by

$$e_{jk} = \frac{1}{\sigma_j} \int \psi_{jk} dB_t^H = \frac{1}{\sigma_j} \sum_{j'k'} w_{j'k'} \int \psi_{jk} v_{j'k'}(t) dt.$$

The near-independence property of e_{jk} can now be shown using the time-scale localization property of wavelets: By normality, the conditional mean $\widehat{e}_{jk} = \mathbb{E}[e_{jk} | e_{j'k'},\ (j',k') \neq (j,k)]$ must lie in the l^2-span of $\{e_{j'k'},\ (j',k') \neq (j,k)\}$: $\widehat{e}_{jk} = \sum_{(j',k')\neq(j,k)} a_{j'k'} e_{j'k'}$. So

[2]The reproducing kernel Hilbert space of dB_t^H consists of $f \in L^2[0,1]$ for which

$$\int_0^1 K_H f(t) f(t) dt < \infty.$$

For standard Brownian motion, the case $H = \frac{1}{2}$, this space is the Cameron-Martin space of Brownian motion: the Sobolev space of absolutely continuous functions f with $f(0) = 0$ and $f' \in L^2[0,1]$.

$$e_{jk} - \widehat{e}_{jk} = -\sum_{j',k'}[\int v_{j'k'}(t) \sum_{j'',k''} \frac{1}{\sigma_{j''}} a_{j''k''}\psi_{j''k''}dt]w_{j'k'} \tag{A.2}$$

$$= -\sum_{j',k'}[\int \psi_{j'k'}(t) \sum_{j'',k''} \frac{1}{\sigma_{j''}} a_{j''k''}v_{j''k''}dt]w_{j'k'}, \tag{A.3}$$

where the second equality follows from the symmetry of the operator $(-\Delta)^{\frac{1}{4}-\frac{H}{2}}$. Using the fact that $\{\psi_{jk}\}$ is an orthonormal basis,

$$\mathrm{Var}(e_{jk} - \widehat{e}_{jk}) = \sum_{j',k'}[\int \psi_{j'k'}(t) \sum_{j'',k''} \frac{1}{\sigma_{j''}} a_{j''k''}v_{j''k''}dt]^2 \tag{A.4}$$

$$= \| \sum_{j'',k''} a_{j''k''} \cdot \frac{1}{\sigma_{j''}} v_{j''k''}\|^2_{L^2} \tag{A.5}$$

$$= c_0 \| \sum_{j'',k''} a^2_{j''k''}\| \tag{A.6}$$

$$> 0. \tag{A.7}$$

This proves Theorem 2.4.5.

Proof of Theorem 2.4.6: Let $f^{(j)}$ be the L^2-projection of f onto the j-th resolution detail subspace, and $R_n^{(j)}(p, q, \zeta, L)$ be the minimax risk of the corresponding Gaussian sequence model with independent noise. By Parseval's equality

$$\mathbb{E}[\|\hat{f}_n - f\|^2] = \sum_j \mathbb{E}[\|\hat{f}_n^{(j)} - f^{(j)}\|^2],$$

and

$$R_n(p, q, \zeta, L) = \sum_j R_n^{(j)}(p, q, \zeta, L).$$

Therefore Theorem 2.4.6 follows immediately from Theorem 2.4.4 and the truncation argument of Lemma A.1.

Proof of Theorem 2.4.8:

Lemma A.2. *Under Assumptions 1 and 5,*

$$nhE[(\frac{1}{n}\sum_{t=1}^{n} K_{h_n}(t - t_0)X_t\epsilon_t)^2] = O_p(1),$$

where $\epsilon_t = \sigma(X_t, t)w_t$.

Proof of Lemma A.2: This can be proven using an argument similar to that from Lemma 3 of [29]).

Proof of Theorem 2.4.8: A similar argument as that in Theorem 2.2.1 shows that, using the same notation,

$$\hat{\kappa}(t_0) - \kappa(t_0) + O_p(h^\rho) = (\frac{1}{n}\sum_{t=1}^{n} K_{h_n}(t-t_0)X_t X_t')^{-1}(\frac{1}{n}\sum_{t=1}^{n} K_{h_n}(t-t_0)X_t\epsilon_t),$$

where $(\frac{1}{n}\sum_{t=1}^{n} K_{h_n}(t-t_0)X_t X_t')^{-1} \in O_p(1)$.

By Lemma A.2,

$$\hat{\kappa}(t_0) - \kappa(t_0) = O_p(h^\rho) + O_p(\frac{1}{\sqrt{nh}}).$$

Therefore

$$\|\hat{\kappa}(t_0) - \kappa(t_0)\|^2 = O_p(h^{2\rho}) + O_p(\frac{1}{nh})$$

and the theorem follows by minimizing with respect to h.

Appendix B

Proofs for Chapter 4

We first fix notation: $\mathbb{R}_+^{m_1 \times m_2}$ denotes the set of $m_1 \times m_2$ matrices with non-negative entries and $(\cdot)^T$ is the transpose operation on matrices. Whenever applicable, $A = [a_{ij}]$ and $A' = [a'_{ij}]$ will denote the adjacency matrices of a network before and after a perturbation with respective maximal eigenvalues λ and λ'. For two $m_1 \times m_2$ matrices A and B, $A - B \geq 0$ means $A - B \in \mathbb{R}_+^{m_1 \times m_2}$.

The following fact will be used repeatedly: Let $A, B \in \mathbb{R}_+^{n \times n}$ with maximal eigenvalues λ_A and λ_B respectively. Then $B \leq A$ implies $\lambda_B \leq \lambda$ with strict inequality if A is irreducible.

Proof of Proposition 4.3.1 Suppose there is a $j \neq i$ such that $\frac{d'_j}{d_j} > \frac{d'_k}{d_k}$ for all $k \neq j$. Let $r_j \in \mathbb{R}_+^{1 \times n}$ be the j-th rows of A and A' (only the i-th rows of A and A' differ). Then

$$\frac{d'_j}{d_j} = \frac{\lambda r_j d'}{\lambda' r_j d} < \frac{r_j d'}{r_j d} = \frac{\sum_{k=1}^n a_{jk} \frac{d'_k}{d_k} d_k}{\sum_{k=1}^n a_{jk} d'_k} < \frac{d'_j}{d_j},$$

which is impossible. □

Proof of Proposition 4.3.2 Let $x \overset{P}{\mapsto} Px$ denote the linear map on $\mathbb{R}^n$ where Px is the vector obtained by setting the i-th entry x_i of x to 0. The assumptions $A'd' = \lambda' d'$ and $d_i = 0$ imply

$$PAPd' = \lambda' Pd'.$$

This implies the maximal eigenvalue λ'' of PAP is $\geq \lambda'$. But $\lambda'' \leq \lambda$. So $\lambda \geq \lambda' \geq \lambda$, i.e. $\lambda = \lambda'$. So

$$A'd' = \lambda d'$$

Since $A, A', A - A' \geq 0$, we have $Ad' = \lambda d'$. By the uniqueness of maximal eigenvector up to scalar multiples, the proposition holds. □

Proof of Proposition 4.3.4 Let $\Delta r \in \mathbb{R}^n_+$ be the additional liability incurred by firm i. By assumption, $(\Delta r)^T d = 0$. If we normalize so that $d_i = d'_i = 1$, then $d \geq d'$ by Theorem Proposition 4.3.1 and a continuity argument. So $(\Delta r)^T d' = 0$ and,

$$Ad' - \lambda' d' = e_i (\Delta r)^T d' = 0$$

i.e. $Ad' = \lambda' d'$. So $\lambda \geq \lambda' \geq \lambda$. Same argument in the previous proof now shows that d and d' must be the same. □

Lemma B.1. *If $T \in \mathbb{R}^{n \times n}_+$ has maximal eigenvalue λ and $\alpha > \lambda$, then $\alpha - T$ is invertible and $(\alpha - T)^{-1} \geq 0$.*

Proof. Since μ is an eigenvalue of T if and only $\alpha - \mu$ is an eigenvalue of $\alpha - T$, $\alpha - T$ is invertible. Without loss of generality, assume $\alpha = 1$, then the hypothesis $1 > \lambda$ implies the series

$$I + T + T^2 + \cdots$$

converges (entry-wise); its limit is necessarily $(1 - T)^{-1}$. Since $T \geq 0$, it is clear that $(1 - T)^{-1} \geq 0$.

□

Lemma B.2. *If invertible matrices T and $S \in \mathbb{R}^{n \times n}$ are such that $T - S \geq 0$, T^{-1}, and $S^{-1} \geq 0$, then $S^{-1} \geq T^{-1}$.*

Proof. We have the general algebraic expression $S^{-1} - T^{-1} = S^{-1}(T - S)T^{-1}$. Since all three matrices on the right-hand side are nonnegative, so is their product. □

Proof of Proposition 4.3.5 Normalize so that $d_i = d'_i = 1$. From $Ad = \lambda d$, we have

$$Pd = (\lambda - PAP)^{-1} PAe_i$$

where P is the projection map from the proof of Theorem 4.4. and e_i is the i-th standard basis vector. Similarly,

$$Pd' = (\lambda' - C)^{-1} w'$$

where $0 \leq C \leq PAP$ and $0 \leq w' \leq PAe_i$, by the hypothesis that $\lambda' \geq \lambda$ and the normalization assumption on the i-th entry of d and d'. By Lemma A.1., $\lambda' - C$ and $\lambda - PAP$ are invertible with nonnegative inverses. Since $\lambda' \geq \lambda$, $\lambda' - C \geq \lambda - PAP$. Lemma A.2 implies

$$(\lambda' - C)^{-1} \leq (\lambda - PAP)^{-1}.$$

So

$$Pd' = (\lambda' - C)^{-1} w' \leq Pd = (\lambda - PAP)^{-1} PAe_i,$$

which proves *(i)*. The proof for *(ii)* is analogous and *(iii)* follows from *(i)* and *(ii)*. □