A PRODIGIUS AMOUNT of financial securities’ trading volume is being generated across the world. These transactions are initiated by traders with heterogeneous motives, are conducted in distant markets of differing platforms, and are subject to a myriad of—sometimes contrasting—trading rules imposed by regulators. All these dimensions weave into a complex nature of financial securities’ trading and these multifarious trades constitute the very fundamental building blocks of the finance industry. It is, therefore, of paramount importance to understand the financial securities market—its functions, development, organization, design, and structure—and so is the aim of this thesis, as well as the future research that might bear its inspiration.

To march afar one always begins with a first step. Bearing the ambition to add to the ultimate understanding of financial securities’ trading, this thesis focuses on how some recent technological advancement have shaped the latest market structure. On the extensive margin, the technological advancement (together with the pro-competition attitude by the regulators) has had the terrain of financial markets trembled: The entry barrier of new market places has been significantly lowered, as witnessed by the sprouting of new high-tech trading venues that fragmented the market share of incumbent markets in the past decade. On the intensive margin, technology improvement has also triggered numerous transformations: Boosts in trading speed have enabled new trading strategies; new patterns from the market data are observed; adapted new players, computerized/algorithmic traders, join and start to dominate the arena; and new battlefields, alternative trading systems (electronic communication networks, crossing networks, dark pools, and the alike), are opened up and developed into different niches of trading needs. The subsequent three chapters of this thesis unveil the details of these recent technological changes, jointly contemplating on some new frictions therein and their implications, yet with respective own emphases on the various measures of market quality.

Chapter 2, based on Menkveld and Yueshen (2014a), is devoted to one of the most dramatic incidents in the history of financial securities trading: the Flash Crash. Both public and proprietary trade data on E-mini (S&P500 future) and SPY (S&P500 ETF) are used. The proprietary dataset allows the identification of the exact trades by a large fundamental seller, who allegedly ignited the nine percent evaporation of the entire U.S. stock market value (Dow Jones Industrial Average) in less than twenty minutes. It is shown that, however, this large seller chose to sell only mindfully and very limitedly during the period when the prices went through a free. Her price impact was magnified by other traders’ aggressive sells with a 300-millisecond delay. As the large seller kept selling (and turned even more aggressive) when the E-mini price bottomed, she reinforced the price pressure and paid excessive cost for transacting a large position. A calibration exercise suggests that her loss during the twenty-minute Flash Crash amounted to a quarter of her annual operating income.

The high-frequency empirical analysis points to the vulnerability of modern financial markets due to possible, unfortunate interaction among the market participants. Such a
crash inflicts a huge transaction cost on the fundamental participants of the market. It seems to be a legitimate, worrying concern that the possibility of incurring prohibitive transaction cost might shake investors’ confidence in the integrity of the current financial market.

Adapted from Menkveld and Yueshen (2014b), chapter 3 looks at the new middlemen who intermediate between end-users in financial securities trading. These middlemen are best thought of as represented by the high-frequency market makers that populate almost all electronic trading platforms nowadays. Compared to the existing literature on middlemen, the insight lies in the multiplicity of such middlemen: Do they interact with each other and, if so, how? What is the economic rationale? More importantly, what does such interaction imply for fundamental investors?

The chapter develops a tractable framework to address the above issues. Through the lens of the model, the effects of middlemen on welfare (allocative efficiency) of the market are scrutinized and identified exactly. In particular, both positive (liquidity provision) and negative (impairing information learning) channels are decomposed from the total marginal effect of an additional middlemen, and the sign of the net effect is pinned down to a parameter that measures how likely the middlemen are able to resell the asset.

It is shown that a fundamental investor might be “confused” by the market activity in that an observed transitory price pressure (due to middlemen multiplicity) could not be distinguished from a permanent price innovation (e.g. revelation of bad quality of the asset). When the former is the case, such inference problem will add to the overall transaction cost of the fundamental investor who inevitably trades on the price pressure and bears the associated risk inefficiently. The model implications appear reminiscent of the narrative of the Flash Crash (as outlined in chapter 2). It is argued that the mechanism discovered by the model could have played a significant role during the Flash Crash.

Chapter 4 is adapted from Yueshen (2014). It explores the optimal limit order submission strategies in an electronic limit order book, with the friction of random latencies between traders’ decision-making (submitting orders) and the decisions’ effectuation (orders being processed by the exchange server). The unknown latencies effectively queue the traders’ limit orders randomly (possibly dependent on traders’ different attributes, e.g., connection speed), and due to the time priority enforced by most of the real-world trading platforms, the associated profitability of the orders varies according to the queue realization.

Recent hardware upgrades, both on the exchange’s servers and on market participants’ end, have magnified the importance of such “queuing uncertainty”. The model developed in chapter 4 helps understand both the nature and the consequences of this friction. The novelty of the model, compared to the bulk of the literature on limit order market, lies in that instead of playing a (perfect information) sequential game, the agents move simultaneously, unable to observe the real-time market status. The model is applied to generate dynamics of liquidity provision, which complements the existing literature on the formation of a stable limit order book; to explain empirically observed phenomena (like so-called
CHAPTER 1. ROAD MAP

ghost/phantom liquidity in the market); and to weigh in on the debate of optimal market
design. It also provides a useful guidance on the empirical works that evaluate how trading
speed (high-frequency traders’ participation, server speed boosts, co-location service, etc.)
affects equilibrium order book depth.

Drawn from independent research papers, these three chapters combine to make a small
step toward a better understanding of financial securities’ trading. It is a mere attempt to
tackle a few of the many outstanding issues arising from the drastic development in the
field. There are yet many unexplored aspects, begging future research to clarify their roles
in resource allocation, asset valuation, information revelation, and so forth.

Such quests lie in the very heart of this thesis. It honors an idealized belief that these
small steps accumulated will push forward the boundary of the literature—the knowledge
base of human beings, enabling a broader horizon with a deeper understanding of the so-
ciety, the economy, and the finance world in particular. Thus is the direction to which this
thesis proudly points.