The Pennsylvania State University
The Graduate School

ESSAYS ON FINANCIAL ECONOMICS AND ECONOMETRICS

A Dissertation in
Economics
by
Shengbo Zhu

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2020

The dissertation of Shengbo Zhu was reviewed and approved by the following:

A. Ronald Gallant
Liberal Arts Professor of Economics
Chair of Committee, Dissertation Advisor

Patrik Guggenberger
Liberal Arts Professor of Economics

Keisuke Hirano
Professor of Economics

Shouyong Shi
Liberal Arts Professor of Economics

Jingzhi Huang
David H. McKinley Professor of Business and Professor of Mathematics

Marc Henry
Professor of Economics
Graduate Program Director

Abstract

In a recent seminal paper, Steve Ross proposed an attractive strategy to extract the physical distribution and risk aversion from just state prices. However, empirical papers that try to use his Recovery Theorem almost all lead to a depressing conclusion: the recovery theorem does not work. Both the state-price matrix and the recovered physical transition matrix are unreasonable and highly sensitive to subjective specifications and constraints. Borovička, Hansen and Scheinkman (2016) proposes a widely-accepted explanation for the empirical failure: according to the Hansen-Scheinkman decomposition established in Hansen and Scheinkman (2009), the assumption about the stochastic discount factor in Ross (2015) is equivalent to arbitrarily setting the martingale component to be 1, which is quite unlikely in reality. In Chapter 1, I argue that in contrast to Borovička, Hansen and Scheinkman (2016), the assumption about the stochastic discount factor in Ross (2015) actually does not set the martingale component in the Hansen-Scheinkman decomposition to be 1. What causes the empirical failure is actually a time-homogeneous state-price matrix, which induces quite restrictive implications on the underlying price process and those restrictions are easily violated in reality. In particular, when the underlying price is used as the state variable or as one component of the state vector, this restriction becomes an eigenvalue equation that contradicts the important eigenvalue equation in Ross (2015), which in this case makes the Recovery Theorem not just empirically implausible, but also logically inconsistent.

Chapter 2 studies the following conceptual question: in what sense is the Fundamental Theorem of Asset Pricing similar to the two-period no-arbitrage theorem (a.k.a., Farkas lemma)? The purpose of studying this question is (1) to study the information that can be extracted from prices of derivatives in a multi-period context, generalizing the result in a two-period case in Breeden and Litzenberger (1978); (2) to find a way to write down explicitly a multi-period arbitrage process, just as a two-period arbitrage can be written down as a vector.

To answer the above conceptual question, I break it down into three more specific questions: (1) How to generalize the concept of states to a multi-period model? (2) How to generalize the concept of state price to a multi-period model? (3) In what sense is a multi-period arbitrage process similar to a two-period arbitrage strategy which is just a vector?

The key to answering those questions is to explicitly describe the probability space on which price processes are defined, especially what "information flow" means. I adopt

the canonical probability space (i.e., the space of all possible paths of some price process) and propose to consider the whole path of as the state variable and the "path prices"(i.e., the equivalent martingale measure) as the analogue of state prices. This chapter discusses how we can recover prices of paths using prices of associated derivative securities and then use them to price other derivatives, which contributes to the literature of implied processes. In addition, it also shows that a multi-period arbitrage process can be reduced to a random vector. The theoretical contribution of this chapter is that it sheds new light on the nature of arbitrage processes and the Fundamental Theorem of Asset Pricing. Practically it provides a general framework to precisely extract the information contained in prices of frequently-traded derivatives and then price other derivatives.

Chapter 3 derives the asymptotic properties of the maximum likelihood estimator and the quasi-maximum likelihood estimator constructed from a Markov hypothesis in the context of a dependent process without making assumptions about the functional form of the likelihood functions. Moreover, this chapter also examines the relation between the two asymptotic distributions and describes the conditions under which the asymptotic variance of the QMLE converges to that of the MLE when more and more lags are used in the construction of the QMLE.

Table of Contents

Acknowledgments

I thank my advisor A. Ronald Gallant for his encouragement and insightful discussions during the time when I was writing this dissertation and his support in the past four years. I also thank Patrik Guggenberger, Jingzhi Huang, Keisuke Hirano, Shouyong Shi for carefully reading the dissertation and insightful feedbacks.

Moreover, it would be impossible to complete this dissertation without understanding and support from my parents. I thank them very much for always providing me with the best education they can provide in the past three decades, supporting me to pursue my an academic career and teaching me what life should be like.

Chapter 1

Why Does the Recovery Theorem Not Work Empirically? An Intrinsic Explanation

1.1 Introduction

It has been known since Breeden and Litzenberg (1978) that in a two-period model, the whole risk-neutral distribution of the price of some asset is implied by prices of European options with different strike prices written on that asset. In a simple multi-period setting such as a Black-Scholes economy, the whole risk-neutral measure is determined by two deterministic parameters, the drift coefficient and the volatility coefficient. The former equals the short interest rate in the economy, while the latter is implied by the price of any European option. So roughly speaking, under some assumptions about the asset price process, the risk-neutral distribution of the process is revealed by prices of derivatives written on that asset.

But when it comes to the physical measure, we have a different story. Again, take the Black-Scholes economy for example. In this case, the whole physical distribution is determined by two deterministic parameters: drift and volatility. Even for this simple case, backing out the whole physical distribution is not an easy task. On the one hand, the volatility is the same under both the physical measure and the risk-neutral measure so that the volatility under the physical measure is directly implied by options prices.

On the other hand, the drift coefficient is changed when the underlying measure changes from the physical measure to the risk-neutral measure and the physical drift coefficient is not mentioned anywhere in the Black-Scholes formula .

If the goal is to know what the physical drift coefficient (or any feature of an asset price process under the physical measure) is, then a natural alternative way is to statistically estimate that coefficient, instead of recovering it from observed prices of associate derivatives, since we can observe a realization of the asset price process. Then it becomes a typical statistical problem, and there is a large literature on it. For example, Chernov and Ghysels (2000) estimates parameters under both measures jointly based on the underlying asset price and a sequence of option prices with different strikes and different maturities, using the Efficient Method of Moments and the reprojection method developed in Gallant and Tauchen (1996) and Gallant and Tauchen (1998).

But is there a way to “recover”, rather than “statistically estimate”, the physical distribution of the underlying asset price process from option prices, just as we do with its risk-neutral counterpart? Ross (2015) provides a positive answer with an elegant method—the Recovery Theorem. The goal of Ross (2015) is quite similar to that of Chernov and Ghysels (2000): both papers try to infer what the dynamics of the underlying asset price look like under both the physical measure and the risk-neutral measure based on option prices. What distinguishes Ross (2015) from other papers which use statistical methods is that Ross(2015) proposes a direct method to calculate, rather than estimate, the dynamics of the underlying asset price under both the risk-neutral measure and the physical measure, with reasonable assumptions.

The two key assumptions are that the underlying asset price process is a time-homogeneous Markov process under both the physical and risk-neutral measure and that the stochastic discount factor (henceforth SDF) has a transition-independent form. Both assumptions appear to be innocuous and widely used. For example, in a typical state-space model the price process is a time-homogeneous Markov process under the physical measure. The Black-Scholes economy is a time-homogeneous Markov process under the risk-neutral measure. In a typical consumption-based asset pricing model with a time-additive utility function, the SDF satisfies the transition-independent assumption. So the assumptions in Ross (2015) are widely used in the literature and we expect good, or at least reasonable empirical results from Ross's Recovery Theorem (henceforth RRT).

However, empirical papers that use the RRT all end up with a negative conclusion: the RRT does not work empirically, and sometimes even produces very unreasonable results.[1] Then the next important and interesting question is: Why? Borovička, Hansen and Scheinkman (2016) (henceforth BHS (2016)) proposes a widely-accepted explanation for the empirical failure: according to the Hansen-Scheinkman decomposition (henceforth H-S decomposition) established in Hansen and Scheinkman (2009) (henceforth HS (2009)), the assumption about the SDF, i.e., the transition-independent assumption in Ross (2015), is equivalent to arbitrarily setting the martingale component to be 1, which is quite unlikely in reality.

In this paper, I re-examine the relationship between the RRT and the H-S decomposition and find that the transition-independent assumption (henceforth T-I assumption) about the SDF actually does not set the martingale component in the H-S decomposition to be 1. Moreover, I find that what causes the empirical failure is likely to be a time-homogeneous state-price matrix, which induces quite restrictive implications on the underlying price process and those restrictions are easily violated in reality. In particular, when the underlying price is used as the state variable or as one component of the state vector, this restriction becomes an eigenvalue equation that contradicts the important eigenvalue equation in Ross (2015), which in this case makes the Recovery Theorem not just empirically implausible, but also logically inconsistent.

The rest of this paper goes as follows: Section 2 surveys related papers related to Ross (2015). Some of them try to generalize the RRT in different directions, and others try to apply the RRT to real data. In Section 3, I briefly review what the H-S decomposition is and then discuss why the T-I assumption does not set the martingale component in the H-S decomposition to be 1. Section 4 discusses implications of a time-homogeneous state-

[1] See Section 2 for a detailed survey of related papers.

price matrix and how it leads to intrinsic inconsistency when the price of the underlying asset is used as the state variable. Section 5 concludes the paper.

1.2 Literature Review

The surprising theorem obtained based on common assumptions in Ross (2015) attracted many other authors to the recovery problem. One natural follow-up problem is how to generalize the RRT from the discrete-time, finite-state context in Ross (2015) to other contexts. Walden (2017) generalizes the RRT to a continuous-time, unbounded univariate diffusion setting, while keeping the transition-independent assumption about the SDF in Ross (2015), and shows a sufficient and necessary condition for recovery to be possible in this setting. Qin and Linetsky (2016) studies the recovery in a more general continuous-time context when the underlying state variable follows a Borel right process, while assuming the SDF is a multiplicative functional of the state process, and shows that recurrence is sufficient for recovery. Carr and Yu (2012) shows that in a continuous-time, bounded univariate diffusion setting, when the numeraire portfolio is also a diffusion driven by the state under both physical and risk-neutral measure, recovery is also possible. Dubynskiy and Goldstein (2013) points out that the boundedness condition plays a key role in enabling recovery in Ross (2015). Schneider and Trojani (2019) generalizes the RRT to a more general framework, although with less information recovered: with only sign restrictions on some covariances between market returns and the pricing kernel, and the risk premium of variance and higher-order moment risks, they characterized the family of pricing kernels that are consistent with these sign restrictions. Jensen et al. (2019) also generalizes Ross (2015) to a large extent: the authors dropped the assumption that the underlying asset price process is a time-homogeneous Markov process under both the physical and the risk-neutral measures, and did not make any probabilistic assumptions. Their recovery strategy relies solely on the form of the algebraic relation between the physical and risk-neutral measures, which is induced by transition-independent SDFs. But as a result of avoiding making assumptions about the probabilistic structure of the underlying process, what is recovered is not some general feature about the process, but transition probabilities of the process at a particular time conditional on a particular state. Dillschneider and Maurer (2019) derives the exact continuous-state analogue of the RRT, where the finite state space is replaced by a continuous state space, state-price matrices are replaced by pricing operators, time could be discrete or continuous, and no other assumptions in Ross (2015) are changed. Tran (2019) generalizes the RRT to a context where the state variable is a diffusion process under both the physical and risk-neutral measure, and the short rate and the risk premium are deterministic functions of the state variable and time, and the SDF has a particular form which is essentially equivalent to the transition-independent assumption in Ross (2015). Under these assumptions a second-order nonlinear ODE is derived that the SDF has to satisfy but uniqueness of the SDF is not guaranteed. Huang and Shaliastovich (2014) relaxes the transition-independent assumption about the SDF and shows a recovery strategy in a recursive-utility framework.

Other papers in this literature test the RRT with real data, and nearly all of them find that the RRT does not work well empirically. For example, Jackwerth and Menner (2018) empirically examines the RRT with prices of options on S&P 500 index. They first followed how Ross did to recover the state-price matrix and got very unrealistic results[2]. They then imposed various kinds of economics constraints in the optimization procedure of recovering the state-price matrix and proceeded to use the RRT to get the physical transition matrix. However, when testing whether observed data was drawn from the recovered physical marginal distribution, all results related to the RRT were rejected while two benchmark models, one of which uses a CRRA utility to obtain the physical transition probabilities from the risk-neutral transition probabilities and the other simply uses historical empirical distribution, passed the test. Audrino et al. (2019) uses a neural-network method to recover the state-price matrix and they found that the recovered SDF is time-inhomogeneous, which directly contradicts the setting of the RRT. Tran and Xia (2018) argues that there is numerical instability caused by the specification

[2]For example, the recovered state-price matrix implies that the risk-free rate varies from -98% to 576%.

of the number of states when implementing the RRT. They emphasized that the numeric instability is ex-ante in the sense that the recovered state-price matrix and also the physical transition matrix are very sensitive to the number of states the underlying state variable can possibly take on, even if the true data-generating process is as what Ross (2015) assumes, data is observed without error but the researcher does not know the true number of states. They also gave a sufficient and necessary condition for recovery results to be consistent with different numbers of states. Dillschneider and Maurer (2019) empirically implements the RRT and finds that the recovered SDF is U-shaped, which contradicts the setting in Ross (2015), and the authors concluded the RRT is misspecified. Yao (2018) also examines the RRT empirically and concludes that it cannot fully recover the physical measure, and interestingly, the recovered physical measure predicts the second and fourth moments better than the first and third moments.

1.3 The Hansen-Scheinkman Decomposition

BHS (2016) offers a theoretical explanation for the empirical failure of the RRT. HS (2009) establishes a remarkable decomposition of the SDF, if the price process is a Markov process under the physical measure and the SDF is a multiplicative functional. In view of this decomposition, BHS (2016) points out that the T-I assumption in Ross (2015) amounts to arbitrarily setting the martingale component in the H-S decomposition to be 1, which loses generality and unlikely to true in practice. Bakshi et al. (2018) uses data on US 30-year treasure bonds to test whether the martingale component is 1 in reality. Unsurprisingly, it is not.

Regarding this criticism about the T-I assumption, Martin and Ross (2019) argues that “the question of whether the hypothesis holds is an empirical one”. And for an empirical question, it is hard to give a definitive answer. More often the only correct answer is “it depends”and it is hard to say what it depends on. It could be datasets, people, algorithms or something as minor as just one observation in a dataset. It is not uncommon in economics research that different researchers who use different data or different statistical techniques come up with contradictory conclusions regarding the same empirical question. As the T-I assumption is concerned, recall that Jackwerth and Menner (2018) uses a model with a CRRA utility function as a benchmark to assess the performance of the RRT, and they found that this benchmark beats all recovery methods based on the RRT. Note that assuming a CRRA utility model implies assuming the transition-independent form of the SDF. So the empirical results in Jackwerth and Menner (2018) seem to suggest that the T-I assumption should not be blamed for the empirical failure of the RRT.

But before proceeding to look for other reasons why the RRT does not work in practice, I would like to re-examine the relationship between the T-I assumption and the H-S decomposition. Does the T-I assumption really imply that the martingale component in the H-S decomposition is arbitrarily set to be 1? Let us first briefly review what the H-S decomposition is. Assume time $t \in \mathbb{R}^+$ is continuous. X_t denotes a time-homogeneous Markov process and let X_t be the state variable that summarizes all kinds of uncertainty in the economy. This means that a random payoff at some time t, or the price of an asset at some time t, can be represented as $\psi(X_t)$, where $\psi(\cdot)$ is some deterministic function defined on the state space. Note that functions that represent prices of the same asset at different times are not necessarily the same. Actually studying how these functions change as the time changes is the focus of HS (2009).

Let M_t denote the SDF. Then for any t, we can define an operator $\mathbb{M}_t$ as follows[3]:

$$\mathbb{M}_t\psi(x) \equiv E(M_t\psi(X_t)|X_0 = x),\ \ \psi(\cdot) \in L$$

where L is a Banach space of real-valued functions defined on the state space. Assume for any t, L is closed with respect to $\mathbb{M}_t$.

Proposition 4.1 in HS (2009) shows that if M_t is a multiplicative functional of X_t, then $\mathbb{M}_t, t \in \mathbb{R}^+$ is a semigroup of operators in L. Let $\mathbb{A}$ denote the extended generator of the semigroup $\mathbb{M}_t$ and let $\phi()$ and ρ denote a principal eigenfunction (i.e., a strictly

[3]Throughout this paper "≡" means "defined as"

positive eigenfunction) and its corresponding eigenvalue of $\mathbb{A}$, respectively. Proposition 6.1 in HS (2009) proves that $exp(-\rho t)M_t\phi(X_t)$ is a (local) martingale. In other words, the SDF M_t can be rewritten as

$$M_t = exp(\rho t)\widehat{M}_t\left[\frac{\phi(X_0)}{\phi(X_t)}\right]$$

where $\widehat{M}_t \equiv exp(-\rho t)M_t[\frac{\phi(X_t)}{\phi(X_0)}]$ is a martingale. The above decomposition says the SDF can be multiplicatively decomposed into a deterministic component $exp(\rho t)$, a martingale component $\widehat{M}_t$ and a transient component $\frac{\phi(X_0)}{\phi(X_t)}$. This decomposition is the central result of HS (2009) and it is referred to as Hansen-Scheinkman decomposition in this paper.

On the other hand, the T-I assumption in Ross (2015) is as follows: the one-period SDF can be represented as

$$\delta\frac{h(X_{t+1})}{h(X_t)}, \quad 0 < \delta < 1$$

where δ is the subjective discount factor and $h(\cdot)$ is some positive function of the state variable. By multiplying the one-period SDF t times, it is easy to see that the corresponding M_t here is $\delta^t\frac{h(X_t)}{h(X_0)}$. Compare it with the H-S decomposition, it appears to be the case that the T-I assumption implicitly assumes the martingale component $\widehat{M}_t$ in the H-S decomposition to be 1, as pointed out by BHS (2016).

But there are several details worth noticing:

(1) The eigenfunction $\phi(\cdot)$ and eigenvalue ρ in the H-S decomposition depend on what the semigroup $\mathbb{M}_t$ looks like. Given how $\mathbb{M}_t$ is defined, what $\mathbb{M}_t$ looks like depends on the SDF process M_t and the transition probability matrix of X_t. As a result, what $\phi(\cdot)$ and ρ look like depend on the transition matrix of X_t.

(2) Under the T-I assumption, the SDF process $M_t = \delta^t\frac{h(X_t)}{h(X_0)}$ in Ross (2015) is a multiplicative functional.

(3) The T-I assumption only specifies the deterministic, functional relationship between the SDF M_t and the state variable X_t. And this relationship has nothing to do with the transition probability matrix of X_t. As a result, the T-I assumption has nothing to do with the transition probability matrix of X_t.

(4) The T-I assumption only says there is "some" positive function $h(\cdot)$ of the state variable X_t, instead of a positive "eigenfunction". Given a function $h(\cdot)$ and the SDF $\delta^t \frac{h(X_t)}{h(X_0)}$, does there exist a transition probability matrix of X_t which generates a semigroup $\mathbb{M}_t$ such that $\delta^t \frac{h(X_t)}{h(X_0)} = exp(\rho t)\widehat{M}_t \left[\frac{\phi(X_0)}{\phi(X_t)}\right]$, where $\phi(\cdot)$ and ρ are a principal eigenfunction and its corresponding eigenvalue of the extended generator of $\mathbb{M}_t$, and $\widehat{M}_t \equiv exp(-\rho t)M_t[\frac{\phi(X_t)}{\phi(X_0)}]$ is a martingale? If the answer is affirmative, then the T-S assumption does not implicitly set the martingale component in the H-S decomposition to be 1.

At first sight, the answer to the above question is "possible", as there is no obvious reason that excludes the existence of such a transition probability matrix. What is interesting is that it turns out that any transition probability matrix works, as long as the H-S decomposition holds with this transition matrix and the given function $h(\cdot)$. To see why, it suffices to note that under the T-I assumption, the SDF process $M_t = \delta^t \frac{h(X_t)}{h(X_0)}$ is a multiplicative functional. Then pick any irreducible transition probability matrix P for X_t. The semigroup $\mathbb{M}_t$ in this case can be written as

$$\mathbb{M}_t \psi \equiv \delta^t D^{-1} P^t D \psi$$

where ψ is a payoff vector, the i-th component of which means the payoff at time t when the state is x_i and $D \equiv diag\Big(h(x_1), h(x_2), \cdots, h(x_n)\Big)$. It is easy to see $\mathbb{M}_t,\ t \in \mathbb{N}$ in this case is a semigroup. So the H-S decomposition applies. That is, if the extended generator of $\mathbb{M}_t$ has strictly positive function ϕ with corresponding eigenvalue ρ, then the SDF has a non-degenerate martingale component, which is $exp(-\rho t)M_t[\frac{\phi(X_t)}{\phi(X_0)}]$, although ϕ may be better represented as a vector.

In summary, the T-I assumption does not implicitly set the martingale component in the H-S decomposition to be 1. On the contrary, the relationship between the T-I assumption and the H-S decomposition is that the T-S assumption, together with an irreducible transition probability matrix, provides a simple example of the H-S decomposition.

1.4 The Intrinsic Inconsistency

1.4.1 Ross's Recovery Theorem

Then what else leads to the empirical failure of the RRT, if not the T-I assumption? Before investigating this problem, let us first recall the setting in Ross (2015) and how the RRT works exactly. The setting in Ross (2015) can be summarized as follows:

- t: time index, assumed to be discrete, $t = 1, 2, \cdots, T$;
- X_t: the state of the world at time t, assumed to be a finite-state, time-homogeneous Markov process, taking values from $\{x_1, x_2, \ldots, x_n\}$
- P: the transition matrix of X_t with dimension $n \times n$;
- p_{ij}: the (i, j)th element of P, represents the probability that tomorrow's state is x_j conditional on today's state is x_i;
- Π: the state-price matrix with dimension $n \times n$, assumed to be time-homogeneous;
- π_{ij}: the (i, j)th element of Π, represents the today's price of tomorrow's 1 dollar in state x_j conditional on today's state is x_i;
- ϕ_{ij}: the SDF, a.k.a. pricing kernel, $\phi_{ij} \equiv \frac{\pi_{ij}}{p_{ij}}$;
- Transition-Independent assumption about the SDF: ϕ_{ij} can be represented as

  $$\phi_{ij} = \delta \frac{h(x_j)}{h(x_i)}, \quad 0 < \delta < 1$$

  where δ is the subjective discount factor and $h(\cdot)$ is some positive function of the state

To keep things as simple as possible, I try to make notations consistent as in the previous section. Next, the RRT can be derived as follows:

Combining the definition of ϕ_{ij} and the T-I assumption, we have $\frac{\pi_{ij}}{p_{ij}} = \delta \frac{h(x_j)}{h(x_i)}$. Rewriting this equation in matrix form leads us to

$$D\Pi = \delta PD \tag{1}$$

where $D \equiv diag\Big(h(x_1), h(x_2), \cdots, h(x_n)\Big)$. Also note that P is a probability transition matrix so that

$$Pe = e \tag{2}$$

where e is a vector with every component equal to 1. From (1) we get $P = \frac{1}{\delta} D\Pi D^{-1}$ and substitute this into (2), we have $\frac{1}{\delta} D\Pi D^{-1} e = e$. Rearranging it gives us

$$\Pi z = \delta z \tag{3}$$

where $z \equiv D^{-1}e$.

Note that (3) is an eigenvalue equation. Now assume Π is a positive matrix, which is a mild assumption since it means that regardless of today's state x_i and tomorrow's state x_j, today's value of tomorrow's 1 dollar is positive. Then by Perron-Frobenius theorem, (3) has one and only one solution[4] that has a positive δ and a positive z(of course z is unique up to a scalar). So both δ and z can be obtained by applying Perron-Frobenius theorem to (3). Also note that from z we immediately get D since $z \equiv D^{-1}e$. Lastly, substituting D and δ into $P = \frac{1}{\delta} D\Pi D^{-1}$ completes the recovery of P, which is the transition matrix. Hence the RRT is proved.

The RRT is undoubtedly a surprising result: it is technically simple, has a elegant form but at the same time reveals a lot of information. It extracts the information about the distribution of the underlying asset price from prices of options written on that asset. No statistical estimation, no model calibration for nuisance parameters, and no controversial assumption about the behavior of an agent in the model. It only needs some standard assumptions which are mostly for simplicity (such as time is discrete, state is

[4]By "solution", I mean a pair of δ and z.

driven by a finite-state, time-homogeneous Markov process), the no-arbitrage assumption which translates into the existence of a positive Π, and a widely-used assumption on the SDF.

1.4.2 A Basic Identity

As mentioned in the Introduction, however, the empirical performance of the RRT is worse than expected. In the rest of this paper I will discuss one reason that leads to its empirical failure. In summary, when the price of the underlying asset is chosen as the state variable [5], there is intrinsic inconsistency within the RRT. Specifically, the setting in Ross (2015) implies another eigenvalue equation which, according to Perron-Frobenius theorem, contradicts (3).

To see why exactly, let us first look at a basic identity. The word "basic" is relative to the RRT, and it means the identity holds regardless of the assumption about the SDF. That is, the basic identity holds no matter whether the T-I assumption holds or not. It is indeed very basic and can be derived as follows: recall that Π is the state-price matrix and π_{ij} is the value of 1 dollar if tomorrow's state is x_j and today's state is x_i. From this definition, we immediately have for any state x_i, $S(x_i, t) = \sum_{j=1}^{n} \pi_{ij} S(x_j, t+1)$, where $S(x_i, t)$ denotes the price of the underlying asset, assumed to be a function of the state variable and time as in Ross (2015). Rewriting it in matrix form, we get the following *Basic Identity* (hereafter BI):

$$\mathbf{S}_t = \Pi \mathbf{S}_{t+1} \tag{4}$$

where $\mathbf{S}_t \equiv [S(x_i, t)]$ is a column vector that collects all possible values of $S(X_t, t)$ when X_t is $x_1, x_2, x_3, \cdots, x_n$. Note that (4) holds as long as Π is a state-price matrix and $S(x_i, t)$ denotes the price of the underlying asset, regardless of assumptions about the SDF.

Remark 1 (A Martingale Interpretation of the BI)**.** *Recall that the Fundamental Theorem of Asset Pricing (henceforth FTAP) says that there is no arbitrage if and only if*

[5] Throughout the paper, "chosen as a state variable" actually means "chosen as a state variable or as a component of the state variable if the state variable is multidimensional"

there exists an equivalent measure under which the stock price discounted by the money market account (hereafter MMA, a.k.a. compounded interest rates) is a martingale. The BI actually is the martingale condition in the current context.

Proof. The equivalence between the BI and the martingale condition can be shown by direct calculation. Denote the MMA at time t by R_t. Note that for any $t, t+1$, if $X_t = x_i$, then R_{t+1}/R_t can be represented as $1/\sum_{j=1}^{n} \pi_{ij}$, i.e., the sum of the i-th row of Π. The risk-neutral martingale condition can be written as $S(x_i, t) = E^* \left(\frac{S(X_{t+1}, t+1)}{R_{t+1}/R_t} \middle| X_t = x_i \right)$. Lastly, note that the risk-neutral transition matrix is just Π with each row divided by the sum of the row, so we can rewrite the above risk-neutral expectation as $S(x_i, t) = E^* \left(\frac{S(X_{t+1}, t+1)}{R_{t+1}/R_t} \middle| X_t = x_i \right) = \sum_{j=1}^{n} \left(\frac{S(x_j, t+1)}{R_{t+1}/R_t} \frac{\pi_{ij}}{\sum_{j=1}^{n} \pi_{ij}} \right) = \sum_{j=1}^{n} \pi_{ij} S(x_j, t+1)$ □

Remark 2 (Pricing Semigroup)**.** *It is interesting to note that how the BI resembles the pricing semigroup in HS (2009): in the current context the semigroup is just* $\Pi^t,\ t \in \mathbb{N}^+$.

It is worth pointing out that what restrictions the BI impose the underlying asset price process: the previous remark implies that in a time-homogeneous Markov context, the state-price matrix alone determines the pricing semigroup. The state-price matrix incorporates all relevant information contained in the SDF and the physical transition matrix for the purpose of pricing. In other words, when a state-price matrix is given, we do not have complete freedom to choose what kind of SDF we want to have in the model. Some SDFs may have implications that are inconsistent with what the state-price matrix implies. And this is what the next subsection is about.

1.4.3 The Intrinsic Inconsistency

So far the intrinsic inconsistency has not occurred yet. Note that so far we have not specified the empirical content of the state variable X_t, but to empirically implement the RRT, this step is necessary. Since the RRT assumes that we know Π, the state variable should be chosen in such a way that Π can be easily obtained. A common practice is to let X_t be the price of a stock and if we observe prices of options on that stock with different strikes and the same expiration date $u > t$, we can back out state prices at time u. This important insight is due to Breeden and Litzenberger (1978).

As far as I know, letting X_t be the price of some stock is the only way of specifying the empirical content of X_t such that backing out Π is feasible, reliable and simple. Of course in principle, as a state variable, X_t can be anything, e.g., volatility of some stock price, consumption level, etc. But the point is that how to back out Π should be simpler and more reliable than statistically estimating P, otherwise the recovery problem makes no sense. Taking this into consideration, choosing X_t to be the price of some stock (or some stock market index) and then using the result in Breeden and Litzenberger (1978) seems to be the only way in the context of the RRT, and this is also how Ross (2015) and all empirical papers following Ross (2015) implement the RRT.

But if we proceed in that way, an intrinsic inconsistency occurs. That is, the BI becomes another eigenvalue equation which contradicts the important eigenvalue equation (3). To see this in detail, first let $S(X_t, t) = X_t$. Denote the possible values of X_t in this case by $s_1, s_2, \cdots, s_n$ instead of previously used $x_1, x_2, \cdots, x_n$. And it is worth pointing out that $s_1, s_2, \cdots, s_n$ are positive real numbers because now X_t denotes the price of a stock. Consequently, $\mathbf{S} \equiv (s_1, s_2, \cdots, s_n)'$ is a positive vector.

What does the BI look like in this context? It suffices to find out what $\mathbf{S}_t$ and $\mathbf{S}_{t+1}$ look like. Take the first element of $\mathbf{S}_t$, $S(s_1, t)$, as an example and all other elements can be checked in the same way. $S(s_1, t)$ denotes the price of the underlying stock when $X_t = s_1$. Because of $S(X_t, t) = X_t$, we obtain $S(s_1, t) = s_1$, and analogously we also have $S(s_i, t) = s_i$ for $i = 2, 3, \cdots, n$. Moreover, these equations also hold for $S(X_{t+1}, t+1)$. So we have $\mathbf{S}_t = \mathbf{S}_{t+1} = \mathbf{S}$. The BI in this case becomes the following *Intrinsic Inconsistency Condition* (hereafter IIC):

$$\mathbf{S} = \Pi \mathbf{S}$$

Theorem 1. *When the price of a stock is chosen as the state variable, the IIC holds and it contradicts* $\Pi z = \delta z$.

Proof. The proof for the IIC has already been shown. It suffices to show the contradiction between the IIC and $\Pi z = \delta z$. By Perron-Frobenius theorem, Π, as a positive matrix, has a unique positive eigenvector (unique up to a scalar) and that eigenvector belongs to the dominant eigenvalue, which is guaranteed to be a positive real number. IIC implies 1 is the dominant eigenvalue of Π, $\mathbf{S}$ is the corresponding positive eigenvector and there should be no other positive eigenvector of Π. But $\Pi z = \delta z$ implies that z is another positive eigenvector and it belongs to δ. Since $0 < \delta < 1$, contradiction occurs. □

Remark 3. *If the underlying stock price $S(X_t, t)$ is a time-homogeneous function of X_t (that is, it has the form $S(X_t)$), the above IIC holds. Choosing the underlying asset price as the state variable makes $S(X_t, t)$ a time-homogeneous function of X_t.*

Remark 4. *Can the contradiction be resolved if δ is allowed to be 1 and let z equal $\mathbf{S}$? It technically works but note that z determines the SDF while $\mathbf{S}$ is just the set of possible values of the price of a stock, which is determined by how we discretize the price interval. If we accept $\delta = 1$ and $z = \mathbf{S}$, then the SDF is determined by how we discretize the price interval, which is quite arbitrary and has nothing to do with data. It's equivalent to arbitrarily picking a SDF instead of letting the data speak.*

Remark 5. *Besides the above inconsistency, the IIC also has another interesting implication that has nothing to do with the SDF. According to the Perron-Frobenius theorem, the dominant eigenvalue is bounded from above by $\max_i \sum_{j=1}^{n} \pi_{ij}$. With 1 as the dominant eigenvalue, $\max_i \sum_{j=1}^{n} \pi_{ij} \geq 1$, which means in some state interest rate is non-positive*[6].

1.5 Conclusion

Asset pricing researchers tend to believe it is almost impossible to extract information about the physical measure from prices of derivatives because even in a Black-Scholes economy, there is only parameter that remains to be known about the physical measure,

[6] As mentioned above, Jackwerth and Menner (2018) does find that row sums of the recovered Π varies from -98% to 576%

which is the drift coefficient. But it is not reflected in option prices. Ross (2015) brings hope to the above question. It has inspired, with its elegant theory and controversial empirical results, lots of asset pricing researchers to explore what kind of information is contained in prices of derivatives. Unfortunately, empirical research based on Ross (2015) does not give us positive results.

It is not uncommon in economic research that good theories are not supported by reality, otherwise there won't be so many "puzzles" in literature. The most reasonable explanation for the empirical failure of a good theory is that some assumptions are simply not realistic, even if they are widely used in academic research. In the literature related to the RRT, a widely-accepted explanation for the empirical failure is from BHS (2016), which says that the T-I assumption arbitrarily sets a martingale component of the SDF to be 1 and this is empirically untrue.

In contrast to other papers in the related literature, this paper investigates the empirical failure of the RRT from a purely theoretical perspective. It first re-examines the relationship between the H-S decomposition and the T-I assumption and finds that the T-I assumption actually does not imply the martingale component to be 1. What causes the empirical failure is rather a time-homogeneous state-price matrix because it alone determines the pricing semigroup, which puts many restrictions on the underlying price process. In other words, a given state-price matrix puts many restrictions on the mapping between the state variable and the underlying asset price process. In particular, when the underlying asset price is used as the state variable to implement the RRT, there are two contradictory eigenvalue equations, one of which is caused by the time-homogeneous Markov EMM and the other is caused by the T-I assumption about the SDF. They make the RRT intrinsically inconsistent.

For future research, an interesting question is what kind of concept, if not state-price matrices, we should use to describe the risk-neutral dynamics of the underlying asset price process for the purpose of extracting information about the underlying price process from prices of derivatives.

Chapter 2

General Implied Processes and Direct Implementation of the Fundamental Theorem of Asset Pricing

2.1 Introduction

The concept of "arbitrage" in asset pricing is as fundamental as "rationality" in economics, and the requirement of "no-arbitrage" is the driving force to price assets just as "utility-maximizing" is the driving force to reach an equilibrium. Pricing theories based only on no-arbitrage is more reliable than those based on additional assumptions such as those about stochastic discount factors, investors' beliefs, utility functions and so on. In a finite-state, two-period model, we have the following no-arbitrage theorem (henceforth TPNA theorem): there is no arbitrage if and only if there exists a strictly positive state-price vector.

The structure of the no-arbitrage pricing problem in a finite-state, two-period model is completely transparent in the sense that no arbitrage is operationally verifiable and in the case that an arbitrage exists, it is implementable. That is, suppose we observe prices of some given assets and we know how the payoffs of those assets in the second period depend on the underlying state. For a new security, if we know how its payoff in the second period depends on the state, then we can: (1) conduct no-arbitrage pricing (i.e., determine its price range to exclude arbitrage); (2) implement an arbitrage (i.e., if the price of a new security is observed and it is not in the price range determined by the previous no-arbitrage pricing analysis, then an arbitrage can be written down explicitly and then implemented).

In other words, no-arbitrage pricing theory in such a two-period, finite-state model is "directly implementable". And as a result, it is practically reliable. Some specific no-arbitrage conditions that are actually used in practice (such as put-call parity) are essentially applications of the no-arbitrage pricing theory in a two-period, finite-state model. The following fact makes the TPNA theorem directly implementable: representations of payoffs of existing securities, state price densities, and arbitrage portfolios are simple in a finite-state, two-period model. Payoffs of existing securities are matrices, a state price density is a vector and an arbitrage is also a vector.

The transparent structure of a two-period, finite-state pricing model also makes it feasible to extract information from prices of derivatives. Given how many kinds of derivatives are being traded and their trading volume, the amount of information contained in prices of derivatives is huge. Together with other assumptions, potential applications based on the implied information include inferring people's risk preference, beliefs about the future movement of the underlying price process, state-price densities of the underlying price, etc. A typical example is Breeden and Litzenberger (1978), which shows how to back out the time-T risk-neutral density of the underlying stock from prices of European call options with different strikes and maturity T. Two features help in this case: (1) a two-period context; (2) we know how payoffs of derivatives at time T depend on the stock price at time T. The finiteness of states is not necessary here.

However, when it comes to the multi-period case, things are different. We still have the multi-period analogue of the TPNA theorem—the Fundamental Theorem of Asset Pricing (henceforth FTAP): there is no arbitrage if and only if there exists an equivalent martingale measure. But we don't have the transparent structure of the multi-period no-arbitrage pricing problem as in the two-period case. That is, if we observe prices some assets and we know how these prices depend on the state variable in each period, we still do not know whether there is arbitrage. Note that here applying the TPNA theorem to each consecutive two periods for each starting state will not work because "arbitrage" in a multi-period model means an adapted, self-financing portfolio process X_t such that $X_1'S_1 = 0$, $X_T'S_T \geq 0$ with probability 1, $X_T'S_T > 0$ with positive probability, where S_t is the price process of existing assets in the market. Applying the TPNA to each two consecutive periods for each starting state can only exclude arbitrages that are state-dependent vectors and live only two periods. It cannot rule out self-financing strategies like the following: if stock i's price is 1 at time 1 and 2 at time 2, buy one stock i and sell the suitable number of stock j at time 3 to make the trading self-financing; close your position at time 5 if stock j's price is 5 at that time, otherwise hold the your portfolio until the last period.

In a multi-period model, it is also not clear what the analog of state-price vectors is. An intuitive idea is to generalize state-price vectors to state-price matrices as in Ross (2015). It also shows that when the underlying asset price is used as the state variable, the state-price matrix can be extracted from prices of a panel of European call options (options with different strikes and different securities). This method can be viewed as a direct generalization of Breeden and Litzenberger (1978). But as shown by Jackwerth and Menner (2018), the state-price matrix obtained in this way is very unreasonable. Tran and Xia (2018) also points out the ex-ante numerical instability of the state-price matrix to how the state space, which is a discretized price interval in this case, is selected. Zhu (2020) analyzes how a state-price matrix puts restrictions on the evolution of the underlying price process and in particular, it points out that using the state-price matrix with the underlying price as the state variable makes Ross (2015) intrinsically inconsistent. It is reasonable to expect that those problems will be mitigated if the underlying asset price is not used as the state variable, but note that the reason why we can recover information from prices of derivatives is that we know explicitly how prices of derivatives depend on the state variable, if the underlying asset price is chosen as the state variable. If it is not the case, relating prices of derivatives to the state variable would be a problem.

This paper tries to build a path-price framework to achieve the following goals: (1) clarify how to represent risk-neutral measures in a finite-state, multi-period model and then how to recover it from prices of relevant derivatives as in Breeden and Litzenberger (1978); (2) make the FTAP "directly implementable" just as the TPNA theorem. These two goals justify the title of this paper. It turns out that a canonical probability space is what is needed: a price process and a risk-neutral measure are defined on the space of all possibles paths of the price process; a "state" is a path of the price process and consequently "state price" means "path price", which is actually the risk-neutral measure on that path. In addition, one advantage of using a canonical probability space is that it comes with a natural filtration. Only with an explicit filtration at hand can we clarify the meaning of "adapted"and the structure of an "arbitrage".

The rest of the paper is organized as follows: Section 2 surveys the literature of implied trees. Models in this literature can be viewed as a prototype of the theory developed in this paper. In other words, the theory in this paper can be viewed as theory of general implied processes instead of "trees". Section 3 talks about the motivation for using a canonical probability space, how the filtration and price processes are defined, how to take conditional expectations and most importantly, how an equivalent martingale measure is represented . Section 4 is devoted to theoretical implications and practical applications, which includes how the essential role of an explicit filtration is revealed, how an equivalent martingale measure is related to a state-price vector, how an arbitrage process is related to a portfolio vector, and a simple application to price path-dependent options, which can also be viewed as a simple example of direct implementation of the FTAP. The last part of Section 4 briefly discusses what the theory in this paper implies about the problem of recovering the physical measure. Section 5 concludes the paper.

2.2 Literature Review

Before heading to the next section, let us first briefly review the literature on implied trees, as research in that literature also aims to recover the risk-neutral measure from prices of derivatives. With a tree structure, recovering the risk-neutral measure is typically reduced to recovering transition probabilities from one node to nodes in the next period so there is no conceptual problem of representing the risk-neutral measure.

The notion of path prices, which is a key concept in this paper, goes back at least to Rubinstein (1994), which belongs to the literature on implied trees. Rubinstein (1994) assumes the underlying asset price evolves like a recombining binomial tree and paths with the same terminal node have equal risk-neutral probability mass [1]. With these two assumptions at hand, both nodes of the tree and (risk-neutral) transition probabilities can be recovered from option prices with different strikes and the same maturity date.

Derman and Kani (1994) also assumes a recombining binomial tree and that a panel of option prices (i.e., prices of options with different strikes and maturity dates) is observed. Then both nodes of the tree and transition probabilities can be recovered. But the positivity of transition probability is not guaranteed, and when a negative transition probability happens, they "override" data in a somewhat arbitrary way. To overcome this problem, Derman et al. (1996) assumes the underlying asset price is a trinomial tree, which produces extra degrees of freedom. But in this case, nodes of the tree are not implied as in the binomial case, but need to be specified exogenously at the beginning. Dupire (1994) also assumes a recombining trinomial tree with exogenously specified nodes. But unlike Derman et al.(1996) that extracts transition probabilities directly from option prices, Dupire (1994) does it in two steps: first extract prices of Arrow-Debreu securities associated with all nodes from a panel of option prices, then use the martingale condition on the underlying asset price to compute transition probabilities.

In contrast to these papers, Britten-Jones and Neuberger (2000) assumes a panel of option prices is observed, the underlying asset price evolves as a trinomial tree[2], but it does not aim at recovering the process completely. It derives a representation of the joint distribution of the underlying asset price in two consecutive periods in terms of option prices, and then proceeds to derive a nonparametric representation for the risk-

[1] In other words, prices of the paths with the same terminal node are equal, which is quite arbitrary

[2] What the authors really made in the paper is a continuity assumption, although in fact it is equivalent to a trinomial tree.

neutral expectation of integrated volatility. To extract a process implied by the panel of option prices, it introduces another finite-state Markov process to model the volatility of the underlying asset price. Derman and Kani (1998) also tries to extract an trinomial tree from a panel of option prices, with complicated volatility dynamics that resembles the HJM condition in fixed income pricing theory. For a more detailed review of these papers, see Jackwerth (1999).

Ross (2015) can also be viewed as part of this literature. It assumes a panel of option prices is observed, the underlying asset price evolves as a time-homogeneous Markov process under the risk-neutral measure, and it shows a way to recover the state-price matrix from the panel of option prices, although the ultimate purpose in Ross (2015) is not to recover the state-price matrix.

How does the path-price framework in this paper compare to those mentioned above? Generally speaking, the path price framework impose fewer assumptions, and recovers less information. That is, on the one hand, it only gives some restrictions about the risk-neutral measure $\mathbb{Q}$ rather than tells us what $\mathbb{Q}(\omega)$ is for a path ω. The only assumptions here are interest rate is 1, time is discrete and prices are discretized. The first assumption is without loss of generality[3] and all the papers mentioned above assume time and price are discrete and finite. On the other hand, the path-price framework does not impose a tree structure so that it does not rule out the possibility of going from one state to another far state ex ante, nor does it restrict the way how the number of possible nodes evolves in time. But note that the tree structure is critical for all the papers mentioned above because strategies in implied-tree papers are sensitive to how many branches the tree has at each time. It does not impose a Markov structure as in Ross (2015), either. In this sense, the theory in this paper can be viewed as "a theory of general implied processes".

But "recovers less information" does not necessarily mean less useful. Being able to extract more information might be simply due to some ad hoc assumptions. For the

[3]Derman and Kani (1994, 1998) and Derman et al. (1996) assume the interest rate is a known constant, which is almost the same as assuming it is 1. Also practically it is more reasonable to let it not depend on the underlying asset price than assume it is driven by the underlying asset price.

purpose of pricing some derivatives based on observed prices of other derivatives, the path-price framework has the following advantages:

(1) It is more flexible because it can use prices of any kind of derivatives as input, and it can produce bounds on prices of any kind of derivatives, as long as the terminal pay off is a random variable defined on the canonical probability space Ω.[4];

(2) It is more reliable because it does not impose ad-hoc tree structure on the risk-neutral dynamics of the underlying asset price (or equivalently, on $\mathbb{Q}$) as mentioned above;

(3) It is computationally simple because the problem of extracting information from observed prices of some derivatives and then using that to price other derivatives can be reduced to a standard linear programming problem. See the toy model in Section 4.3 for details.

(4) Besides, as far as I know, the path-price framework is the first no-arbitrage pricing framework that explicitly models the filtration (i.e., the information flow). Other papers in the existing literature either do not mention it at all, or as in Britten-Jones and Neuberger (2000), the filtration is mentioned but does not play a role. Admittedly, the term "information available up to time t" is abstract and vague, and the necessity of explicitly modeling it is not apparent. Therefore, there has not been previous research that tries to model [5] it. However, for the purpose of extracting $\mathbb{Q}$ and then conducting no-arbitrage pricing in a multi-period model, clarifying the content of information is important both in theory and in practice. Note that two important objects in the FTAP, EMM and arbitrage process, are defined with respect to some underlying filtration. See Section 4.1 for more detailed discussion.

[4]That is, the terminal pay off of the derivative depend only on the realized path of the underlying asset price

[5]There are many generals ways to measure the amount of information, such as using the notion of entropy, but "modeling the information up to time t" is not just measuring the amount of information.

2.3 A Path-Price Framework

2.3.1 A Hint From the Proof of the FTAP

In the Introduction we mentioned that the multi-period no-arbitrage pricing problem is quite different from the two-period no-arbitrage pricing problem in terms of the clarity of their structures. Mathematically the latter is isomorphic to Farkas lemma while it is not clear what the former is isomorphic to. But there is one analogy which we believe must be true: the FTAP as a whole should be the generalization of the TPNA theorem as a whole. With this belief in mind, the problem of making the structure of the FTAP transparent is reduced to clarifying the correspondence between components of the FTAP and the TPNA theorem.

A natural idea is to trace the FTAP back to its first formal proof. Harrison and Kreps (1979) and Harrison and Pliska (1981) first formally proved the FTAP[6] in a discrete-time setting and a continuous-time setting, respectively. Although some proofs there are constructive, all objects, including prices, portfolio processes, and equivalent martingale measures (henceforth EMM), on defined on an underlying probability space Ω. For mathematical proofs, objects defined on abstract probability spaces are sufficient. But when the empirical meaning of Ω is not stated explicitly, the FTAP is not implementable. In other words, there is still no way to represent explicitly how prices depend on the underlying state, how an EMM is defined on the state space, and how to implement an arbitrage portfolio process.

2.3.2 The Underlying Path Space

Now the problem is reduced to clarifying the empirical content of the probability space Ω. As long as we clarify the empirical content of Ω and its elements $\omega \in \Omega$, we can simply call an ω a state and $\mathbb{Q}(\omega)$ the price of that state, where $\mathbb{Q}$ is an EMM defined on Ω. A general and convenient way of explicitly constructing the probability space

[6] Depending on the precise mathematical meanings of “no-arbitrage" and “equivalent martingale measure", the FTAP has slightly different versions. But the difference is pure technicality.

when studying stochastic processes is to let Ω be the "canonical probability space"[7], i.e., the space of all possible paths of the underlying asset price. In this case, the empirical content of some $\omega \in \Omega$ is clear: it is just one possible path of the underlying asset price. This subsection sets up the environment.

Let t denote time and take values in $\mathbb{T} \equiv \{1, 2, 3, \cdots, T\}$. Let K_t be a finite subset of $\mathbb{R}_+$, which denotes the set of possible values that the underlying asset price can take on at time t. The set of possible paths for the underlying asset is denoted by $\Omega \equiv \times_{i=1}^{T} K_i$. Thus an element $\omega \in \Omega$ denotes one possible path of the underlying asset, and $\omega_t \in K_t$ denotes the value of the path ω at time t. Let the price of the underlying asset be denoted by S_t. S_t is a random variable, which is a mapping from Ω and take on values in K_t. Given that ω defined as above, S_t is actually a coordinate projection: $S_t(\omega) \equiv \omega_t$[8].

The canonical space comes with a natural and explicit information flow, which is probably the most useful feature of this framework. Having an explicit information flow is important, although the importance is often ignored. There are at least two reasons for its importance: firstly, the FTAP says no-arbitrage is equivalent to the existence of an EMM, but whether a process is martingale or not depends on the information flow; secondly, the definition of an arbitrage portfolio also depends on the information flow since it is required to be an adapted, self-financed portfolio process. Without an explicit information flow, even if we are told there is no EMM so that an arbitrage portfolio process exists, we still cannot find that process because we do not know whether a process is adapted. In other words, an explicit information flow gives us a mathematically rigorous and empirically meaningful definition of the statement "conditional all the information available so far".

The information flow on the canonical space can be defined as the σ-algebra flow generated by the process S_t. It can also be defined in the following explicit way: let $[s_{t_1} s_{t_2} \cdots s_{t_n}], 1 \leq t_1 < \cdots < t_n \leq T, s_{t_i} \in K_i, i = 1, 2, 3, \cdots, n$ denote the set of ω such that $\omega_{t_i} = s_{t_i},\ i = 1, 2, 3, \cdots, n$. For any $t \in \mathbb{T}$, define $\mathcal{F}_t$, which represents all the information available at time t, as the σ-algebra generated by all sets of the

[7]See, for example, Kallenberg (1997), Chapter 7

[8]S_t as a stochastic process is called a canonical process.

form $[s_1 s_2 \cdots s_t]$, $(s_1, s_2, \cdots, s_t) \in \times_{i=1}^{t} K_i$. Now the empirical content of the statement "conditional on $\mathcal{F}_t$" means we know which path we are on up to time t, or equivalently, we know our path is in some set $[s_1 s_2 \cdots s_t]$, but do not further know which subset $[s_1 s_2 \cdots s_t s_{t+1}] \subset [s_1 s_2 \cdots s_t]$ our path is in. Moreover, the statement that "a process X_t defined on Ω is adapted" means for any t, any two paths $\omega^1, \omega^2 \in \Omega$, if $\omega_i^1 = \omega_i^2, i = 1, 2, \cdots, t$, $X_t(\omega^1) = X_t(\omega^2)$. Adaptedness is actually a pretty mild regularity condition for a stochastic process. In our case, it just means prices of derivatives at time t are formed based on only the information up to time t.

Let $\mathbb{Q}$ and R_t denote an EMM defined on Ω and the MMA at time t, respectively. Now we have completed the construction of the underlying probability space:

$$\left(\Omega, (\mathcal{F}_t)_{t=1}^{T}, \mathbb{Q}\right)$$

. Just as in the two-period case where the risk-neutral probability mass on a time-t state discounted by R_t is called the price of that state, $\mathbb{Q}(\omega)/R_T$ can be called the price of the path ω given that in our context a state actually means a path.

To simplify things a bit, assume interest rate is 1 throughout the rest of the paper. It is actually without loss of generality because we can simply reinterpret the paths in Ω as discounted paths and prices of derivative securities as discounted prices. This reinterpretation does affect us much given what we want to explore. Note that for the purpose of recovering a general implied process, the focus here is on the market that consists of the underlying asset and derivative securities written on the underlying asset, whose pay off are determined by the behavior of the underlying asset price S_t. We are mainly studying what information about S_t we can recover from price of those derivatives. Interest rate is an exogenous process, since it is not determined by the behavior of S_t. Completely modeling interest rate would require extending the probability space Ω to $\Omega \times$ the path space of interest rate, which causes significant and unnecessary complexity.

2.3.2.1 The Structure of Conditioning

According to the risk-neutral pricing theory, asset prices are conditional expectations of their future prices or pay offs under the risk-neutral measure. Given an EMM and a derivative security, the price of that derivative, as a stochastic process, is implicitly pinned down. But the problem is how to find an explicit expression for that stochastic process. Almost all kinds of derivative pricing problems are essentially problems of solving for analytical expressions of conditional risk-neutral expectations and it is no different here.

Fortunately with all the structures on the underlying probability space, we have explicit expressions for $\mathcal{F}_t$-conditional expectations of any random variable defined on $\left(\Omega, (\mathcal{F}_t)_{t=1}^T, \mathbb{Q}\right)$. Note that due to the finiteness of paths, any random variable is integrable and can be represented as linear combinations of indicator functions of subsets of Ω, so it suffices to solve for explicit expressions for indicator functions.

Before heading to the general formula for conditional expectations, let us introduce a new notation: $\mathcal{P}_{123\cdots t}: \ \Omega \to \mathcal{F}_t$, $\omega \mapsto [\omega_1\omega_2\cdots\omega_t]$. $\mathcal{P}_{123\cdots t}$ is an operator that maps a path ω, to the set of paths $[\omega_1\omega_2\cdots\omega_t]$, every element of which has the same history up to time t as that of ω. A more intuitive interpretation is that $\mathcal{P}_{123\cdots t}$ reduces the information contained in a whole path to that contained in the 0-to-t segment of that path.

Theorem 2. *For any $A \subseteq \Omega$ and any $t \in \mathbb{T}$, we have*

$$E\left(\mathbb{1}_A|\mathcal{F}_t\right)(\omega) = \frac{\mathbb{Q}\left(A \cap \mathcal{P}_{123\cdots t}(\omega)\right)}{\mathbb{Q}\left(\mathcal{P}_{123\cdots t}(\omega)\right)}, \text{ for any } \omega \in \Omega, \tag{5}$$

Proof. According to the the definition of conditional expectations [9], it suffices to show $\frac{\mathbb{Q}(A\cap\mathcal{P}_{123\cdots t}(\omega))}{\mathbb{Q}(\mathcal{P}_{123\cdots t}(\omega))}$ is $\mathcal{F}_t$-measurable and verify for any $B \in \mathcal{F}_t$,

$$\int_B \mathbb{1}_A d\omega = \int_B \frac{\mathbb{Q}\left(A \cap \mathcal{P}_{123\cdots t}(\omega)\right)}{\mathbb{Q}\left(\mathcal{P}_{123\cdots t}(\omega)\right)} d\omega \tag{6}$$

[9] See, for example, Section 5.1 in Durrett (2010)

By construction, the $\mathcal{F}_t$-measurability is obvious. Note that any subset $A \subseteq \Omega$ is a pairwise-disjoint union of single-path sets and that any $B \in \mathcal{F}_t$ is a pairwise-disjoint union of sets of the form $[s_1 s_2 \cdots s_t]$, where $s_i \in K_i, i = 1, 2, \cdots, t$. So without loss of generality, we can assume A is a singleton $\{\omega\}$ for some $\omega \in \Omega$ and B has the form $[s_1 s_2 \cdots s_t], s_i \in K_i, i = 1, 2, \cdots, t$. Then we have

$$\text{both LHS and RHS of (6)} = \begin{cases} \frac{\mathbb{Q}(\{\omega\})}{\mathbb{Q}([s_1 s_2 \cdots s_t])}, & \text{if } \omega_i = s_i, i = 1, 2, \cdots, t \\ 0, & \text{otherwise} \end{cases}$$

□

Remark 6. *Despite that (5) appears to lack empirical relevance, it actually is a very general asset pricing formula in our context. It can be viewed as a generalization of the Black-Scholes formula in the following sense: the Black-Scholes formula gives us the time-t price of terminal pay off* $max(S_T - K, 0)$ *as a function of time t and* S_t*, and (5) gives us the time-t price of terminal pay off* $\mathbb{1}_A$ *as a function of time t and realized path of the underlying asset price up to time t. In contrast to the Black-Scholes formula, the price process given by (5) is not necessarily a Markov process.*

Remark 7. *The non-Markov property is in contrast to other models in the literature such as in Rubinstein (1994), Dupire (1994) and Derman and Kani(1994), where conditional probability means the transitional probability from today's state to tomorrow's state. Here the conditional probability means given the history up to today, the probability of reaching some state tomorrow.*

Example 1 (Price processes of Arrow-Debreu securities)**.** *As a simple example, let us use (5) to find out how the price of an Arrow-Debreu security evolves. Since a state in our context means a path, an Arrow-Debreu security written on a particular path* ω^0 *is a security that pays 1 dollar if the realized path is* ω^0*, and pays 0 otherwise. Denote the price process of such an Arrow-Debreu security by* $AD[\omega^0](t, \omega)$[10]*. Substituting* $A = \{\omega^0\}$

[10] $AD[\omega^0]$ should be viewed as a whole to represent a process, and since it is a process, it has two arguments, ω and t.

into (5), we get

$$AD[\omega^0](t,\omega) = \begin{cases} \frac{\mathbb{Q}(\{\omega^0\})}{\mathbb{Q}([\omega_1\omega_2\cdots\omega_t])}, & \text{if } \omega_i^0 = \omega_i, i = 1,2,\cdots,t \\ 0, & \text{otherwise} \end{cases}$$

The above price process for $AD[\omega^0]$ is actually quite intuitive. It basically says that when the realized part of the path $\Big($i.e., $(\omega_1,\omega_2,\cdots,\omega_t)\Big)$ until time t coincides with that of ω^0, then the price of $AD[\omega^0]$ progresses as the ratio of the risk-neutral probability of ω^0 to that of the sum of all possible paths that could still occur. If not, then the security $AD[\omega^0]$ vanishes, simply because it is impossible that the actual realized path is ω^0, and consequently at time T the one dollar will not be paid.

2.3.2.2 The Martingale Condition and Regularity Assumptions

According to (5), nearly all kinds of derivatives, including European options, Asian options, barrier options[11], and the underlying asset itself, can also be priced since their terminal payoff can be represented as linear combinations of indicator functions. But the underlying asset is a bit special, because its pricing problem is actually solved when it is defined as a coordinate projection: $S_t(\omega) \equiv \omega_t$. So now regarding the time-t pricing of the payoff $S_{t+1}(\omega)$ for any $1 \le t \le T-1$, we have two ways: one is through definition, time-t price is $S_t(\omega) \equiv \omega_t$; the other way is using (5). Note that $S_{t+1}(\omega) = \sum\limits_{s_{t+1}\in K_{t+1}} s_{t+1}\mathbb{1}_{[s_{t+1}]}(\omega)$, then by applying (5) we obtain

$$E(S_{t+1}|\mathcal{F}_t)(\omega) = \sum_{s_{t+1}\in K_{t+1}} s_{t+1}\frac{\mathbb{Q}([s_{t+1}]\cap\mathcal{P}_{123\cdots t}(\omega))}{\mathbb{Q}(\mathcal{P}_{123\cdots t}(\omega))}$$

These two ways should be equivalent, otherwise there is intrinsic inconsistency. In other words, $S_t(\omega) = E(S_{t+1}|\mathcal{F}_t)(\omega)$ for any $\omega \in \Omega$. Since this should hold for any $1 \le t \le T-1$, it actually means S_t is a martingale. We now show this equivalence condition can be reduced to an ex-ante condition on the EMM $\mathbb{Q}$:

[11] But not American options since their payoff depend on not only the realized path of the underlying asset, but also on the exercise time

Proposition 1 (Martingale Condition). *For $S_t(\omega) = E(S_{t+1}|\mathcal{F}_t)(\omega)$ to hold for any $\omega \in \Omega$ and any $1 \leq t \leq T-1$, it is necessary and sufficient that the following Martingale Condition on $\mathbb{Q}$ holds:*

For any $1 \leq t \leq T-1$ and any $(s_1, s_2, \cdots, s_t) \in \times_{i=1}^{t} K_i$, we have

$$s_t \mathbb{Q}([s_1 s_2 \cdots s_t]) = \sum_{s_{t+1} \in K_{t+1}} s_{t+1} \mathbb{Q}([s_1 s_2 \cdots s_{t+1}])$$

Proof. First notice that both mappings are $\mathcal{F}_t$-measurable[12], so it suffices to check the above equality for all ω that have different history up to t. And by doing that, we arrive naturally at the above Martingale Condition. □

It is worth pointing out that the similarity between the above Martingale Condition (henceforth MC) and the IIC in Chapter 1. They are both the result of two things: the martingale requirement in the FTAP and using the underlying asset price as the state variable[13]. They are both ex-ante restrictions on $\mathbb{Q}$[14]. The MC is actually the generalization of the IIC: when we impose assumptions that $K_1 = K_2 = \cdots = K_T$ and that S_t under $\mathbb{Q}$ is a time-homogeneous Markov process[15], the MC reduces to the IIC.

A natural follow-up question is how many restrictions does the MC impose on $\mathbb{Q}$? Is it possible that the MC over-identifies $\mathbb{Q}$? Denote the cardinality of K_i by $|K_i|$, $1 \leq i \leq T$ and let $|K_1| = 1$, which means we know today's price of the underlying asset. Notice that for any $1 \leq t \leq T-1$ and any $(s_1, s_2, \cdots, s_t) \in \times_{i=1}^{t} K_i$, there is a restriction. So in total there are $\sum_{t=1}^{T-1} \prod_{i=1}^{t} |K_i|$ restrictions. On the other hand, there are $\prod_{i=1}^{T} |K_i|$ paths in total and all path prices should sum to 1 ($\sum_{\omega \in \Omega} \mathbb{Q}(\omega) = 1$), since $\mathbb{Q}$ is a probability measure. To avoid over-identification, we need $\prod_{i=1}^{T} |K_i| - 1 - \sum_{t=1}^{T-1} \prod_{i=1}^{t} |K_i|$ to be positive. The following assumption gives a very natural sufficient condition:

[12]That is, if any two paths ω^1 and ω^2 have same values up to time t, then $S_t(\omega^1) = S_t(\omega^2)$, $E(S_{t+1}|\mathcal{F}_t)(\omega^1) = E(S_{t+1}|\mathcal{F}_t)(\omega^2)$

[13]The fact that the pricing problem for S_t is solved when $S_t \equiv \omega_t$ is a consequence of the assumption that ω means a possible path of S_t

[14]Here "ex-ante" means "caused by the setting the model, instead of told by data"

[15]This means all kinds of conditional probabilities in the MC can be summarized by a transition matrix

Assumption 1 (Uncertainty Assumption)**.** *For any* $2 \le i \le T$, $|K_i| \ge 2$ *and there exists some* i_0 *such that* $|K_{i_0}| \ge 3$.

Proposition 2. *If the Uncertainty Assumption holds, then* $\prod_{i=1}^{T} |K_i| - 1 - \sum_{t=1}^{T-1} \prod_{i=1}^{t} |K_i| > 0$.

Proof.

$$\frac{\sum_{t=1}^{T-1} \prod_{i=1}^{t} |K_i| + 1}{\prod_{i=1}^{T} |K_i|} = \frac{1}{\prod_{i=1}^{T} |K_i|} + \frac{1}{\prod_{i=2}^{T} |K_i|} + \frac{1}{\prod_{i=3}^{T} |K_i|} + \cdots + \frac{1}{|K_T|}$$
$$< \frac{1}{2^{T-1}} + \frac{1}{2^{T-1}} + \frac{1}{2^{T-2}} + \cdots + \frac{1}{2} = 1$$

□

Remark 8. $|K_i| \ge 2$ *means there is uncertainty about the price of the underlying asset. Otherwise if for some* j, $|K_j| = 1$, *then it means at time* j *the price has only one possible value, which amounts to be known ex ante.* $|K_{i_0}| \ge 3$ *rules out the case where* $|K_2| = |K_3| = \cdots = |K_T| = 2$ *and* $\prod_{i=1}^{T} |K_i| = \sum_{t=1}^{T-1} \prod_{i=1}^{t} |K_i| + 1$, *which means at any time and with any kind of realized history, the uncertainty is always a two-point distribution. Note that we have at least two assets in our model (the underlying asset and cash), so the market is complete and* $\mathbb{Q}$ *is exactly identified before observing any data.*

Actually ruling out over-identification is not enough. For this path space to be a good platform to study the information implied by prices of derivatives, we need to rule out a more general condition: ex-ante conditional monotonicity. That is, for any time $1 \le t \le T-1$ and any $s_t \in K_t$, the price space in the next period, K_{t+1}, should allow the underlying asset price to go both up and down[16]. Otherwise if we know when $S_t = s_t^0$ for some $s_t^0 \in K_t$, the price is definitely going to go up (or down) at time $t+1$, then we have an arbitrage strategy: buy (or sell) the underlying asset if the underlying asset price at time t is s_t, otherwise do nothing. Following this strategy, our terminal pay off is non-negative and could be positive with positive probability. It can be easily seen that this kind of ex-ante arbitrage, which is caused by the setting of the price space, is ruled out if the following assumption holds:

[16]Note that if $|K_{t+1}| = 1$, this cannot be the case so $|K_i| \ge 2$ is a necessary condition for ruling out ex-ant conditional monotonicity.

Assumption 2 (Expanding Price Spaces). *Let* $\underline{s}_t \equiv$ *the minimum in* K_t *or* $\overline{s}_t \equiv$ *the maximum in* K_t*,* $t = 1, 2, 3, \cdots, T$*. We assume* $\underline{s}_1 > \underline{s}_2 > \cdots > \underline{s}_T >$ *and* $\overline{s}_1 < \overline{s}_2 < \cdots < \overline{s}_T$*.*

The above also rules out the case where "$\overline{s}_i = \overline{s}_{i+1}$"(or "$\underline{s}_i = \underline{s}_{i+1}$"). In this knife-edge case, the MC implies that the measure $\mathbb{Q}$ will concentrate on paths $[\overline{s}_i\overline{s}_{i+1}]$ (or $[\underline{s}_i\underline{s}_{i+1}]$) and for any $s'_{i+1} \in K_{i+1}$ and $s'_{i+1} \neq \overline{s}_{i+1}$ (or $\underline{s}_{i+1}$), $\mathbb{Q}([\overline{s}_i s'_{i+1}])$ (or $\mathbb{Q}([\underline{s}_i s'_{i+1}])$) $= 0$. Intuitively it means that under $\mathbb{Q}$, when the underlying asset price hit the maximum (or minimum), it is going to stay there forever. This kind of ex-ante degeneracy should also be ruled out.

It is interesting to compare the setting in Ross (2015) and the current path space. The time-homogeneous, Markov process in Ross (2015) implies the path space there does not satisfy Assumption 2. That is, in Ross (2015) we have "$\overline{s}_i = \overline{s}_{i+1}$" and "$\underline{s}_i = \underline{s}_{i+1}$". Then the MC in this case reduces to the Intrinsic Inconsistency Condition (henceforth IIC) as pointed out in Chapter 1, which implies the above ex-ante degeneracy. More specifically, suppose the maximum of $\mathbf{S}$ is s_i, so the i-th row of the IIC is that $s_i = \sum_{j=1}^{n} \pi_{ij} s_j$. Since s_i is the maximum element in $\mathbf{S}$ and $0 \leq \pi_{ij} \leq 1$, for $s_i = \sum_{j=1}^{n} \pi_{ij} s_j$ to hold, it has to be the case that $\pi_{ii} = 1$, $\pi_{ij} = 0$ when $j \neq i$. The same argument works for the minimum of $\mathbf{S}$ as well.

At this moment we have constructed the underlying risk-neutral probability space $\left(\Omega, (\mathcal{F}_t)_{t=1}^{T}, \mathbb{Q}\right)$, talked about the general formula for conditional expectations, and imposed some ex-ante restrictions on Ω and $\mathbb{Q}$. The general formula for conditional expectations can be used for two things: on the one hand, when $\mathbb{Q}$ is known, price a derivative at any time t with any realized history $(s_1, s_2, \cdots, s_t) \in \times_{i=1}^{t} K_i$; on the other hand, when $\mathbb{Q}$ is unknown but prices of some derivatives are observed, we can recover some aspects of $\mathbb{Q}$ by combining the general formula for conditional expectations and the observed prices of those derivatives. The partially-recovered $\mathbb{Q}$ can be further used to price other exotic derivatives, or at least give bounds. This is probably the main use of this path price framework in practice and the next section has a toy example to illustrate how these problems can be reduced to linear programming problems, as well as some new and interesting insights when ω is regarded as a path of the underlying asset price, into the role of the information flow, the relation between the one-period no-arbitrage condition and the FTAP, and the nature of a self-financing strategy.

2.4 Theoretical Implications And an Application to Relative Pricing

This section discusses two kinds of theoretical implications of the path-price framework and then shows how it can be used to do relative pricing in a toy model. The two theoretical implications shed new light on concepts such as filtration, EMM and arbitrage processes and the relative pricing method provides a new, general and simple way to extract and use information contained in prices of derivative securities. More specifically, we are going to first talk about the role of the filtration in the context of no-arbitrage pricing, the relation between the FTAP and the TPNA theorem and how to use the path-price framework to reduce the problem of relative pricing to a linear programming problem.

2.4.1 On the Role of the Filtration

This subsection is about the implication of having an explicit filtration in the path-price framework. What if the filtration is left unspecified? As mentioned in the previous section, the filtration is important since EMM and arbitrage portfolio processes have to be defined with respect to some filtration. Consequently, the problem of whether there is an arbitrage or not (or equivalently, whether exists an EMM) is not well-defined until the underlying filtration is given. Roughly speaking, a finer filtration will allow more self-financing portfolio processes to be adapted hence become arbitrages, and as a result, puts more restrictions on an equivalent measure to be an EMM. So a no-arbitrage market might contain arbitrages if the filtration is enlarged.

Without explicitly specifying the filtration, there is no way to make the meaning of "equivalent martingale measure" and "adapted self-financing portfolio process" precise and consequently, it is not clear how to write down what the FTAP tells us precisely. In other words, it is impossible to unleash the power of the FTAP exactly without explicitly specifying the filtration. Note that the FTAP is a theorem of type "Nonexistence of object A is equivalent to existence of object B". To use this kind of theorems exactly, precise definitions of object A and B are important. A basic logic principle for this kind of theorems is that if definitions of A and B are changed (usually for simplicity) but the theorem still holds, then a relaxed definition of A corresponds to a restricted definition of B, and a more restricted definition of A corresponds to a relaxed definition of B. In either case, the power of the original theorem is not fully unleashed.

One typical case where a relaxed definition of A corresponds to a restricted definition of B is restricting the EMM to be a Markovian EMM. That is, an investor assumes the risk-neutral measure is not only a martingale measure but also a Markov measure [17], just as in Ross (2015). This Markovian-style thinking assumes we know more about the EMM than what the FTAP actually tells us. In other word, the existence of a Markov EMM does not only exclude the existence of arbitrages, but something more. When used with empirical data, it restricts the class of potential EMMs so much that there is little room left to let data speak about what the EMM looks like, hence proceeding in this way ends up with very poor data fitting, just as in a typical over-identified situation. The path-price framework is a good context to reflect to what extent the Markov EMM puts the model at the risk of over-identification by simply counting unknowns and knowns: there are $\prod_{i=1}^{T} |K_i|$ path prices, the Martingale Condition imposes $1 + \sum_{t=1}^{T-1} \prod_{i=1}^{t} |K_i|$ restrictions, and the Markov assumption will additionally induce another $\sum_{t=1}^{T-1} \prod_{i=1}^{t} |K_i|$ restrictions. It can be seen that an extra Markov assumption does not necessarily lead to over-identification, but it does induce significantly more restrictions (as many as what the Martingale Condition induces). When there are other assumptions (such as an assumption about the form of the SDF), a Markov assumption also increases the risk of contradicting those assumptions, which is what happens in Ross (2015).

[17] Or equivalently, the underlying asset price is a martingale and also a Markov process

One typical case where a restricted definition of A corresponds to a relaxed definition of B is thinking of the "adapted and self-financing portfolio processes" in a Markovian way, as in Duffie and Garman (1989). That is, an investor thinks of a multi-period model as many two-period models glued together without information progression through those two-period models, an arbitrage portfolio process is represented by a deterministic vector (i.e., not contingent on observed information as there is no information to be contingent on) which are trivial examples of adapted processes, just as in a two-period model and consequently, "no arbitrage process" in this case actually means for any two periods, there is no arbitrage in this two-period sub-model. The feature of "being adapted" is missing. As a result it relaxed the definition of EMM, i.e., enlarges the class of potential EMMs hence does not fully unleash the power of the FTAP[18]. And it is no wonder that Duffie and Garman (1989) proves there always exists a Markovian EMM since the class of potential EMMs is enlarged.

Compared to the above two Markovian-style theories, the path-price framework can be viewed as in the middle: neither does it assume more about the EMMs than what the FTAP says, nor does it restrict the class of arbitrage processes. The Martingale Condition captures precisely what the FTAP says about EMMs and given how the filtration is defined, "adapted" means "contingent on the observed path of the underlying asset price at any time". So under the path-price framework, the FTAP is used in an exact way and its power is fully unleashed.

2.4.2 On the FTAP as a Generalization of the TPNA Theorem

This subsection is about how the path-price framework illustrates the analogy between a state-price vector in a two-period model and an EMM in a multi-period model, and between a arbitrage portfolio vector in a two-period model and an arbitrage portfolio process in a multi-period model. In other words, this subsection illustrates in what sense the FTAP can be viewed as a generalization of the TPNA theorem.

[18]That is, we could have known more than what the FTAP tells us in this case, if the usual definition of arbitrage was adopted.

2.4.2.1 An EMM as a Family of History-Dependent State-Price Vector

To illustrate the first kind of similarity, note that the martingale requirement in the FTAP in general can be written as $E^{\mathbb{Q}}(S_{t+1}|\mathcal{F}_t) = S_t$ $\mathbb{Q}$-a.s.. Under the path-price framework, this $\mathbb{Q}$-a.s. equality has a precise and explicit meaning. That is, according to Proposition 1, the martingale condition can be represented as

$$s_t\mathbb{Q}([s_1s_2\cdots s_t]) = \sum_{s_{t+1}\in K_{t+1}} s_{t+1}\mathbb{Q}([s_1s_2\cdots s_{t+1}])$$

for any $1 \le t \le T-1$ and any $(s_1, s_2, \cdots, s_t) \in \times_{i=1}^{t} K_i$. If $\mathbb{Q}([s_1s_2\cdots s_t]) = 0$, then the equality trivially holds since $\mathbb{Q}([s_1s_2\cdots s_{t+1}]) \le \mathbb{Q}([s_1s_2\cdots s_t]) = 0$. If $\mathbb{Q}([s_1s_2\cdots s_t]) > 0$, then the equality can be rewritten as

$$s_t = \sum_{s_{t+1}\in K_{t+1}} s_{t+1}\frac{\mathbb{Q}([s_1s_2\cdots s_{t+1}])}{\mathbb{Q}([s_1s_2\cdots s_t])}$$

If we see $\frac{\mathbb{Q}([s_1s_2\cdots s_{t+1}])}{\mathbb{Q}([s_1s_2\cdots s_t])}$ as the state price of s_{t+1} conditional on $S_i = s_i, i = 1, 2, \cdots, t$, then the existence of an EMM implies for any $1 \le t \le T-1$, there exists a state-price vector $q_t(\omega)$ that is $\mathcal{F}_t$-measurable and strictly positive $\mathbb{P}$-a.s. [19] and satisfies $S_t = S'_{t+1}q_t$ $\mathbb{P}$-a.s.. In other words, the EMM can be seen as a family of history-depedent, strictly positive state-price vector.

2.4.2.2 An Arbitrage Portfolio Process as a History-Dependent Portfolio Vector

In general, an arbitrage in the two-period case can be represented as a deterministic vector while an arbitrage in a multi-period model is defined to be a self-financing, adapted portfolio process. Apparently the latter is much more complicated than the former and it is not obvious that in what sense the latter is a generalization of the former. However we can show that in our current context (i.e., a finite-state, finite-time model with finite assets), a multi-period arbitrage always implies a history-dependent vector, which is also

[19] Note that the $\mathbb{P}$-a.s. strict positivity of q_t corresponds to the equivalence of $\mathbb{Q}$.

an arbitrage.

More specifically, let $\left(\Omega, \mathcal{F}, (\mathcal{F})_{t=1}^{T}, \mathbb{P}\right)$ be a filtered probability space, where Ω is a finite set (not necessarily a set of paths). S_t is a positive, d-dimensional stochastic process defined on Ω and it denotes price processes of all assets in the market. Assume there exists a risk-less asset in the market. Let X_t be a d-dimensional stochastic process and it denotes a portfolio process. Let $W_t \equiv X_t' S_t$ be the wealth process of the portfolio process X_t.

The timing in the multi-period model is as follows: first, S_1 is observed, then the investor chooses X_1, and then the price process randomly fluctuates to S_2, then X_2 is chosen $\cdots$. The key thing is that X_t is chosen after S_t is observed, and this is basically what "adapted" means. Moreover, the self-financing condition requires that

$$X_{t-1}' S_t = X_t' S_t$$

for $t = 2, 3, \cdots, T$, which means the wealth does not change before and after a transaction.

Now we can prove the following general result about the reduction of an arbitrage in a multi-period model:

Theorem 3. *If $X_1, X_2, \cdots, X_T$ is an arbitrage (i.e., $X_t, t = 1, 2, \cdots, T$ is self-financing, and satisfies $X_1' S_1 = 0$, $X_T' S_T \geq 0$ with probability 1, $X_T' S_T > 0$ with positive probability), then at least one of the following two statements holds:*

(i) $X_1, X_2, \cdots, X_{T-1}$ is also an arbitrage for the period of time 1 to time $T-1$;

(ii) There exist a $\mathcal{F}_{T-1}$-measurable set B, a d-dimensional vector b and a positive constant c such that $\mathbb{1}_{B \times \{T-1\}}(\omega, t)(b, c)$ is an arbitrage portfolio process, where the $(d+1)$-dimensional vector (b, c) means buy portfolio b and buy the risk-less asset of value c.

Proof. To prove the theorem, if suffices to show when $X_1, X_2, \cdots, X_{T-1}$ is not an arbitrage for the period of time 1 to time $T-1$ (i.e., when $W_{T-1} = 0$ with probability

1 or $\{W_{T-1} < 0\}$ has positive probability), (ii) holds. First note that because of the self-financing condition, the wealth process $W_t \equiv X_t' S_t,\ 2 \leq t \leq T$ ($W_1 \equiv X_1' S_1$ is always 0) can be represented recursively as $W_t = W_{t-1} + X_{t-1}'(S_t - S_{t-1})$. And also recall that the probability space is finite, so X_{T-1}' can be represented as $\sum_{i=1}^{J} \mathbb{1}_{B_i} b_i$, where $B_i,\ i = 1, 2, \cdots, J$ are disjoint $\mathcal{F}_{T-1}$-measurable sets with positive probability, form a partition of Ω, and $b_i,\ i = 1, 2, \cdots, J$ are distinct d-dimensional vectors.

If $W_{T-1} = 0$ with probability 1, then since $W_T = X_{T-1}'(S_T - S_{T-1}) \geq 0$ with probability 1 and > 0 with positive probability, there must exist some $k \in \{1, 2, \cdots, J\}$ such that $\mathbb{1}_{B_k} b_k (S_T - S_{T-1}) \geq 0$ with probability 1 and positive with positive probability. This means the following strategy is also an arbitrage: from time 1 to time $T-2$, do nothing; at time $T-1$, if B_k happens, buy portfolio b_k; else do nothing. This portfolio process can be represented as $\mathbb{1}_{B_k \times \{T-1\}}(\omega, t)(b_k, 0)$. It is obviously adapted, and for self-financing, it suffices to note that $X_{T-1}' S_{T-1} = 0$ implies $\mathbb{1}_{B_k \times \{T-1\}}(\omega, T-1) b_k' S_{T-1} = 0$.

If $\{W_{T-1} < 0\}$ has positive probability, then there must exist some B_k such that $\{W_{T-1} < 0\} \cap B_k \neq \emptyset$. Let $B \equiv \{W_{T-1} < 0\} \cap B_k$ and note that B is $\mathcal{F}_{T-1}$-measurable and has positive probability. Since $W_T = W_{T-1} + X_{T-1}'(S_T - S_{T-1}) \geq 0$ with probability 1 and > 0 with positive probability, $\mathbb{1}_B b_k (S_T - S_{T-1}) \geq 0$ with probability 1 and positive with positive probability as well. This means the following strategy is also an arbitrage: from time 1 to time $T-2$, do nothing; at time $T-1$, if B happens, buy portfolio b_k and buy the risk-less asset of value $c \equiv |b_k' S_{T-1}|$; else do nothing. This portfolio process can be represented as $\mathbb{1}_{B \times \{T-1\}}(\omega, t)(b_k, c)$. It is obviously adapted since B is $\mathcal{F}_{T-1}$-measurable and for self-financing, it suffices to note that $W_{T-1} \equiv X_{T-1}' S_{T-1} < 0$ on B implies that $\mathbb{1}_{B \times \{T-1\}}(\omega, T-1) b_k' S_{T-1} < 0$, which further implies $\mathbb{1}_{B \times \{T-1\}}(\omega, T-1) b_k' S_{T-1} + c = 0$. □

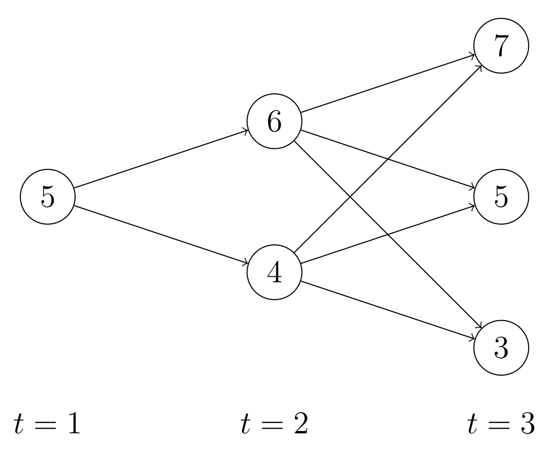
Theorem 3 illustrates in what sense a self-financing, adapted arbitrage portfolio process is a generalization of an arbitrage vector: the former can be reduced to a history-dependent vector (or a deterministic vector, which can be viewed as a special contingent vector). As far as I know, this result is new.

Actually Theorem 3 is not directly related to the path-price framework, but it is expected based on the analogy between an EMM and state-price vectors, which is revealed by the path-price framework. Also, Theorem 3 alone does not make a history-dependent vector implementable. The path-price framework specifies what the statement "$\mathcal{F}_{T-1}$-measurable set B happens" means and makes Theorem 3 implementable. In this regard we have the following simple proposition:

Proposition 3. *When the probability space Ω is taken to be the set of paths of some underlying asset price process, and suppose the market consists of the underlying asset and all derivative securities whose terminal payoff depend on the realized path of the underlying asset price, then any arbitrage can be reduced to a "wait-and-buy" strategy, i.e., wait for some particular pattern*[20] *of the underlying asset price to happen, then buy some predetermined portfolio, if that pattern does not occur, do nothing.*

2.4.3 Relative Pricing in the Path-Price Framework

In this subsection let us use the path-price framework to price exotic derivatives. It turns out that all such problems can be reduced to linear programming problems. Consider the following simple toy model as illustrated by the following path space:



[20]This particular pattern corresponds to the $\mathcal{F}_{T-1}$-measurable set B in Theorem 3

Let $q_{5ij} \equiv \mathbb{Q}([5ij])$ denote the price of the path $5 \longrightarrow i \longrightarrow j$, where $i = 4,6$ and $j = 7,5,3$. Then according to the Martingale Condition, we have

$$5 = 6 \times (q_{563} + q_{565} + q_{567}) + 4 \times (q_{543} + q_{545} + q_{547}) \tag{7}$$

$$6 \times (q_{563} + q_{565} + q_{567}) = 7 \times q_{567} + 5 \times q_{565} + 3 \times q_{563} \tag{8}$$

$$4 \times (q_{543} + q_{545} + q_{547}) = 7 \times q_{547} + 5 \times q_{545} + 3 \times q_{543} \tag{9}$$

Of course, we have additionally that

$$\sum_{i=4,6\ j=3,5,7} q_{5ij} = 1 \tag{10}$$

$$q_{5ij} \geq 0,\ i = 4,6, \quad j = 3,5,7$$

Before using these equations to price complicated derivatives, let us first test whether they rule out some common arbitrage opportunities. Recall that the market under consideration consists of the underlying asset, cash with interest rate 1, and all kinds of derivatives whose terminal pay off depends on the path of the underlying asset price. Typical arbitrage strategies in this market are portfolios of the underlying asset, cash, European call and put options. The following proposition shows that put-call paritys[21] for any put-call pair is captured by equations (7)-(10).

Proposition 4. *If there exists an EMM $\mathbb{Q}$ in the current toy model and $\mathbb{Q}$ satisfies (7)-(10), then for any European put-call pair, their corresponding put-call parity holds.*

Proof. Denote the time 1 price of an arbitrary (European) call and put option by $C(K,T)$ and $P(K,T)$, respectively, where K is the strike price and T is the maturity date. Note that in the current toy model, when $T = 2,\ K = 4,6$ and when $T = 3,\ K = 3,5,7$. So there are 6 put-call parities. Let's take $T = 3, K = 7$ for example. In this case the put-call parity is

$$C(7,3) + 7 \times 1 = P(7,3) + 5$$

[21]Put-call parity means for any European call and put option with the same strike price K and maturity date T, $C(K,T) + KP(T) = P(K,T) + S$, where $P(T)$ is the current bond price with maturity date T and S is the current underlying asset price.

, where 1 is the price of a unit zero-coupon bond that matures at time 3. Since 7 is the highest possible price at $T = 3$, we trivially have $C(7, 3) = 0$. Under the path-price framework, $P(7, 3) = 2 \times (q_{565} + q_{545}) + 4 \times (q_{563} + q_{543})$, so the put-call parity can be represented in terms of path prices as follows:

$$2 \times (q_{565} + q_{545}) + 4 \times (q_{563} + q_{543}) = 2 \tag{11}$$

. Now the question is whether this is captured by the Martingale Condition together with the normalization and the non-negativity restriction, i.e., whether (7)–(10) imply (11).

It is indeed the case. From (7) and (10) we know that

$$q_{563} + q_{565} + q_{567} = q_{543} + q_{545} + q_{547} = \frac{1}{2} \tag{12}$$

. Combining (12) and (8) gives us

$$q_{565} + 2q_{563} = \frac{1}{4} \tag{13}$$

, and similarly, from (12) and (9) we can derive

$$q_{545} + 2q_{543} = \frac{3}{4} \tag{14}$$

Substituting (13) and (14) into the LHS of (11) proves that (11) holds.

For other combinations of K and T, the proof can be done in the same way. □

Now consider the pricing problem of an Asian call option $A(3, 3)$, which means the strike price $K = 3$, the maturity date $T = 3$, and the terminal pay off is the average realized price minus the strike price or 0, whichever is greater. Apparent it is a path-dependent option. Under the path-price framework,

$$A(3, 3) = q_{543} + \frac{5}{3}q_{545} + \frac{7}{3}q_{547} + \frac{5}{3}q_{563} + \frac{7}{3}q_{565} + 3q_{567}$$

The pricing problem of this Asian option can be reduced to the following linear programming problem:

$$\max_{\substack{q_{5ij}\geq 0\\ i=4,6\ j=3,5,7}} \text{ and } \min_{\substack{q_{5ij}\geq 0\\ i=4,6\ j=3,5,7}} A(3,3) = q_{543} + \frac{5}{3}q_{545} + \frac{7}{3}q_{547} + \frac{5}{3}q_{563} + \frac{7}{3}q_{565} + 3q_{567}$$

$$\text{subject to } (7) - (10)$$

By solving it we know that $2 \leq A(3,3) \leq 2$, which means this asian option happens to be priced accurately by the MC. Analogously we can show that for another Asian option with a higher strike price, say $A(5,3)$, we have $\frac{1}{3} \leq A(5,3) \leq \frac{3}{8}$.

When prices of some derivative securities are observed, then we can represent these observed prices in terms of path prices, add them to the set of constraints, and follow the above steps. The path-price framework can incorporate information contained in prices of nearly all kinds of derivatives (particularly convenient for dealing with path-dependent options), even observed at different times, as long as they can be represented in terms of path prices.

What is worth pointing out is that when prices of some derivatives are observed at the same time, and then used as additional constraints in a linear programming to price another derivative, if the feasible set is empty, then it signals that there is an arbitrage in the market consisting of the underlying asset and all those derivatives the observed prices of which are used as constraints. This arbitrage opportunity is practically reliable because in this case we are essentially in a two-period model if we "vectorize" the multi-period model with each path as a component of the vector, and an arbitrage in a two-period model is practically reliable. Note that we do not make any assumption about the behavior of any price process under the physical distribution. The only assumption here is discretization, but this can be mitigated by discretizing the price interval and time interval as finely as possible.

Admittedly, computation is a problem since the dimension of the vector increases exponentially with the number of periods T. Another interesting and realistic computational problem is when the feasible set is empty, how to find an arbitrage? The analogue of this problem in a two-period model is when there is no state-price vector, how to find an arbitrage, which is equivalent to how to find a separating hyperplane that separates a point and a convex cone. In the multi-period case, after "vectorizing" the model, an arbitrage is not just a sequence of separating hyperplanes, but also self-financing and adapted. These two are essentially new requirements that do not show up in the two-period case.

2.4.4 On the Recovery of the Physical Measure

Lastly, let us look at what can be said about recovering the physical measure from the risk-neutral measure in the context of the path-price framework. First of all, a physical measure in the case is also a measure defined on the space of paths, just like the risk-neutral measure. Secondly, if we look back at previous sections, it seems that the physical measure is no where mentioned except that it is equivalent to the risk-neutral measure, which in the context means $\mathbb{P}(\omega) = 0$ if and only if $\mathbb{Q}(\omega)$ for any $\omega \in \Omega$. As a consequence we seem to arrive at the conclusion that the risk-neutral measure does not convey any information about the physical measure except for the fact that these two measures are equivalent, if no additional assumptions are present. Is this still the case if we leave the underlying probability space unspecified? Note that using the path space as the underlying probability space does not introduce any essential assumption so the irrelevance between the risk-neutral measure and the physical measure should still hold.

To get an intuitive understanding of the reason, let us first use the two-period, finite-state case as a simplified example. It is quite clear that the risk-neutral measure (or the state price vector in this case), has nothing to do with the physical measure. Note that the state-price vector is determined by the payoff matrix, regardless of what the physical distribution is[22]. Analogously, in the multi-period, finite-state case, the risk-neutral measure (or the family of path prices in this case), has nothing to do with the physical measure. It is determined by asset price processes which are mappings from $\Omega \times \mathbb{T} \to \mathbb{R}_+^d$, where d is the number of assets that exist in the market. In short, the

[22]As long as every state has positive physical probability.

risk-neutral measure is determined by some mappings defined on $\Omega \times \mathbb{T}$. The physical measure, on the other hand, is something defined on Ω and there is no ex-ante relation between what these mappings look like and what the physical measure looks like so that there is no ex-ante relation between the risk-neutral measure and the physical measure. In other words, the risk-neutral measures depends on the functional dependence, instead of the probabilistic dependence, between price processes. Actually "risk-neutral measure" is just an interpretation of the object that is called a risk-neutral measure. The object is interpreted in this way because it is a real-valued, $\mathbb{P}$-a.s. positive and $\mathbb{P}$-integrable function defined on Ω so as a result, it can be viewed as a probability measure after rescaling.

Of course, when asset price processes, viewed as mappings from $\Omega \times \mathbb{T} \to \mathbb{R}^d_+$, are linked with the physical measure by some assumption, e.g., they are Itô processes, then the risk-neutral measure is also linked to the physical measure in some way besides that they are equivalent. A typical example is that the risk-neutral volatility is the same as that under the physical measure when the asset price processes are Itô processes.

2.5 Conclusion

In a two-period, finite-state model, Breeden and Litzenberger (1978) shows how to recover the future risk-neutral distribution of an underlying stock from prices of options written on that stock. Also in such a model, we can determine whether there exists arbitrage in the market and if it does exist, how to implement it. But when it comes to a multi-period, finite-state model, neither of the above thing is possible, if no additional assumption on the underlying asset price process is made.

This paper shows that to make the FTAP as transparent as the TPNA theorem, the key is to construct the underlying probability space in an explicit way. The canonical probability space, i.e., the set of paths of the underlying asset price process, is a good choice for this task: it comes with the natural filtration, $\omega \in \Omega$ means a possible realization of the price process, which is explicit, and the underlying price process is just the coordinate projection. As a result, risk-neutral measures are defined on the space of paths, and the risk-neutral probability mass of a path can be naturally called the price of that path, which can be viewed as the generalization of the concept of state prices in the two-period case. Moreover, the path-price framework constructed in this way also sheds new light on the role of the information flow, the analogy of an arbitrage in the two-period case and an arbitrage process in the multi-period case. It also provides a new and easy way to extract information from some derivative securities to price other derivatives. Finally, as a byproduct, since the path-price framework provides a clear view of the fundamental mathematical structure underlying any no-arbitrage pricing context, it reveals that the risk-neutral measure is not related to the physical measure except that they are equivalent measures, if no additional assumption is imposed.

Chapter 3

Markovian Quasi-Maximum Likelihood Estimation

3.1 Introduction

Quasi-maximum likelihood estimation has been widely studied in the past several decades, due to its intuitive plausibility, relative regularity and easiness to compute, interesting properties of likelihood functions previously discovered, and perhaps most importantly, the hope of replicating various optimal properties of maximum likelihood estimation. A quasi-likelihood function is constructed based on an assumption about the functional form of the true likelihood function. Although it is often the case (if not always) that the assumption is wrong, we can nevertheless derive meaningful properties about the resulting estimator. One of the most commonly used quasi-likelihood function is the Gaussian likelihood, which is constructed based on the belief that the distribution of some underlying random variable is Gaussian, whose true distribution is actually unknown. For such an example, see Hall and Yao (2003). For a systematic treatment on the quasi-likelihood estimation, see Heyde (1997).

As described above, what the quasi-likelihood function and the resulting estimator look like is a consequence of the assumption based on which it is constructed. Different such assumptions will certainly lead to quite different estimators, and how to select this assumption is often more like art. For example, when it comes to estimating a parameter in the context of a dependent process, where there is no specific assumption about the dependence structure such as linearity, then Gaussian assumption is not applicable.

This paper considers this case and construct the quasi-likelihood function from a Markov perspective. This idea is first seen in Gallant and Long (1997), where they show the EMM/SNP estimator is asymptotically equivalent to the quasi-maximum likelihood estimator that is obtained by replacing each transition density in the correct likelihood by a transition density on L lags. But such a QMLE has more good properties than just being asymptotically equivalent to the EMM/SNP estimator. For example, in the context this paper is focused on, it can be represented as an average of a stationary and strong mixing sample, while the MLE cannot be represented in this way. Consequently, to derive the asymptotic properties of the QMLE, only moment conditions are needed while in the case of MLE, we need near epoch dependence conditions. The plan for the rest of the paper is as follows: in Section 2, asymptotic results are stated and proved; in Section 3, the focus is on the relation between the asymptotic distributions when more and more lags are used when constructing the quasi-likelihood function.

3.2 Maximum likelihood Estimation and Markovian Quasi-Maximum Likelihood Estimation

Suppose there is a stationary sequence $\{Y_t\}_{t\in\mathbb{Z}}, Y_t \in R$, and its distribution is parametrized by a parameter ρ, which is also the parameter of interest with the true value being ρ_0. For simplicity, assume $\rho \in \Lambda \subset \mathbb{R}$, where Λ is compact. Denote the density function of $Y_t, Y_{t-1}, \ldots, Y_{t-L}$ by $p_{t,t-L}(Y_t, Y_{t-1}, \ldots, Y_{t-L}, \rho), t \in \mathbb{Z}, L \in \mathbb{Z}_+$. In other words, the functional form of the density function of an arbitrary finite set of the sequence is known up to the parameter ρ. For the simplicity of notation, throughout the paper, random arguments of functions are omitted when no confusion is caused, for example, $p_{t,t-L}(Y_t, Y_{t-1}, \ldots, Y_{t-L}, \rho)$ will be written as $p_{t,t-L}(\rho)$. Let $f_{t,t-L}(\rho)$ denote the conditional density of Y_t conditional on $Y_{t-1}, Y_{t-2}, \ldots, Y_{t-L}$, then obviously we have $f_{t,t-L}(\rho) = p_{t,t-L}(\rho)/p_{t-1,t-L}(\rho)$ when $L \geq 1$.

Suppose the observed sample starts from some time, and without loss of generality, suppose it starts from time 0. The MLE can be defined as follows:

$$\hat{\rho}_n \equiv \arg\max_{\rho\in\Lambda} G_n(\rho)$$

where "≡" means "defined as" and $G_n(\rho) \equiv \frac{1}{n}\sum_{t=1}^{n} \log f_{t,0}(\rho)$. [1]

Note that the functional form of $f_{t,0}(\rho)$ is different for different t, since the number of arguments is increasing with t. This causes troubles both for its computation and deriving its asymptotic results. However, a Markovian condition would get rid of this phenomenon immediately. Suppose that $\{Y_t\}_{t=-\infty}^{\infty}$ is not only stationary but also Markovian, then for $\forall t \in \mathbb{Z}_+$, $f_{t,0}(\rho)$ has the same functional form due to stationarity and every $f_{t,0}(\rho)$ has only two random arguments Y_t, Y_{t-1}. Moreover, $\{f_{t,0}(\rho)\}_{t=1}^{\infty}$ is also a stationary sequence. These observations suggest constructing a Markovian quasi-likelihood function and naturally, there is a corresponding quasi-maximum likelihood estimator:

$$\tilde{\rho}_n \equiv \arg\max_{\rho\in\Lambda} Q_n(\rho)$$

where $Q_n(\rho) \equiv \frac{1}{n}\sum_{t=L}^{n} \log f_{t,t-L}(\rho)$ [2], $L \in \mathbb{Z}_+$ is a constant, which means the number of lags that $f_{t,t-L}(\rho)$ is conditional on. Here we pretend that the sequence $\{Y_t\}_{t=-\infty}^{\infty}$ satisfies that for $\forall t \in \mathbb{Z}$, $\forall L \in \mathbb{Z}_+$, Y_t is independent of $(Y_{t-L-1}, Y_{t-L-2}\dots)$ conditional on $(Y_{t-1}, Y_{t-2}, \dots, Y_{t-L})$. Note that this condition is equivalent to the usual definition of a Markov process when $L = 1$, and strictly weaker than that when $L > 1$.

As can be seen from the above, both the MLE and the QMLE considered in this paper will be of general form, i.e., there is no functional form assumption regarding density

[1] This is not exactly the log-likelihood of the whole sample. If we want to use the exact log-likelihood of the whole sample, then we should write the criterion function from $t = 0$, i.e., $G_n(\rho) \equiv \frac{1}{n}\sum_{t=1}^{n} \log f_{t,0}(\rho) + p_{0,0}(\rho)$. But it can be proved that under the conditions that are used in this section, the consistency of one estimator implies the other and they have the same asymptotic variance.

[2] As in the MLE case, this is also not exactly the log-likelihood of the whole sample. If we want to use the exact log-likelihood of the whole sample, then we should include the unconditional density of $Y_0, Y_1, \dots, Y_L - 1$, and write the criterion function as $Q_n(\rho) \equiv \frac{1}{n}\sum_{t=L}^{n} \log f_{t,t-L}(\rho) + \log p_{t-L-1,0}(\rho)$. But under the conditions used below, the consistency of one estimator implies that of the other and they have the same asymptotic variance

functions. The price of this generality is that some high-level complicated conditions are needed, e.g., stochastic equicontinuity and near epoch dependence. The proofs in this section are actually quite standard when stochastic euqicontinuity and near epoch dependence are assumed. For the simplicity of exposition, I will just assume these two high level assumptions when they are needed, instead of stating long, cumbersome primitive conditions. There is a vast literature on primitive conditions for these two assumptions, for example, see Newey (1991) and Andrews (1994) for primitive conditions for stochastic equicontinutiy, Chapter 7, Gallant (1987) and Chapter 17, Davidson (1994) for primitive conditions for near epoch dependence.

Before heading to the asymptotics, some regularity conditions and a mixing condition should be assumed to avoid stating them repeatedly as they are needed for the asymptotics of both MLE and the Markovian QMLE. Both of them are assumed throughout the paper.

Regularity Conditions

(i) ρ_0 lies in the interior of $\Lambda \subset \mathbb{R}$, which is compact; $E(\log p_{t,t-L}(\rho)) < \infty$ is continuous on Λ and ρ_0 is the unique maximizer of $E(\log p_{t,t-L}(\rho)), \forall t \in \mathbb{Z}, \forall L \in \mathbb{Z}_+$;

(ii) Interchangeability of differentiation and integration: for $\forall t \in \mathbb{Z}, L \in \mathbb{Z}_+$, $\log p_{t,t-L}(\rho)$ is twice continuously differentiable, and

$$int_{\mathbb{R}^{\mathbb{L}}} \frac{\mathrm{d}p_{t,t-L}(y_t,\ldots,y_{t-L+1},\rho)}{\mathrm{d}\rho}\mathrm{d}y_t\cdots\mathrm{d}y_{t-L+1}\Big|_{\rho=\rho_0} = \frac{\mathrm{d}}{\mathrm{d}\rho}\int_{\mathbb{R}^L} p_{t,t-L}(y_t,\ldots,y_{t-L+1},\rho)\mathrm{d}y_t\cdots\mathrm{d}y_{t-L+1}\Big|_{\rho=\rho_0}$$

$$\int_{\mathbb{R}^{\mathbb{L}}} \frac{\mathrm{d}^2 p_{t,t-L}(y_t,\ldots,y_{t-L+1},\rho)}{\mathrm{d}\rho^2}\mathrm{d}y_t\cdots\mathrm{d}y_{t-L+1}\Big|_{\rho=\rho_0} = \frac{\mathrm{d}}{\mathrm{d}\rho}\int_{\mathbb{R}^L} \frac{\mathrm{d}}{\mathrm{d}\rho} p_{t,t-L}(y_t,\ldots,y_{t-L+1},\rho)\mathrm{d}y_t\cdots\mathrm{d}y_{t-L+1}\Big|_{\rho=\rho_0}$$

Mixing Condition

Let $\mathcal{F}_s^t \equiv \sigma(Y_t, Y_{t-1}, \ldots, Y_s)$, the σ-algebra generated by $Y_t, Y_{t-1}, \ldots, Y_s, -\infty \leq s \leq t \leq \infty$. The process $\{Y_t\}_{t=-\infty}^{\infty}$ is strong mixing of size $\frac{-4r}{r-4}$ for some $r > 4$. It might be worth mentioning that $\frac{-4r}{r-4} < -4$ when $r > 4$.

3.2.1 Asymptotic Properties of MLE

The asymptotics of the MLE will rely heavily on the concepts of near epoch dependence and stochastic equicontinuity.

Theorem 4. *Assume Regularity Conditions, Mixing Condition and the following conditions hold:*

1. *For* $\forall \rho \in \Lambda$, $\{\log f_{t,0}(\rho) - E(\log f_{t,0}(\rho))\}_{t=1}^{\infty}$ *is near epoch dependent*[3] *of size* $-\frac{2(r-2)}{r-4}$*;*
2. *For* $\forall \rho \in \Lambda, \sum_{t=1}^{\infty} \frac{\|\log f_{t,0}(\rho) - E(\log f_{t,0}(\rho)\|_r}{t^2} < \infty$ *where* $\|X\|_r \equiv (E(|X|^r))^{\frac{1}{r}}$*;*
3. $\overline{G}_n(\rho) \equiv \frac{1}{n}\sum_{t=1}^{n} E(\log f_{t,0}(\rho)), n = 1, 2, 3, \ldots$ *is equicontinuous and* $G_n(\rho), n = 1, 2, 3, \ldots$ *is stochastically equicontinuous*[4] *on* Λ *.*

Then $\hat{\rho}_n \xrightarrow{p} \rho_0$.

Proof. A standard argument[5] is adopted, i.e., first prove the uniform convergence of $G_n(\rho)$ to $\overline{G}_n(\rho)$ and then consistency follows from the Regularity Conditions. By Theorem 2.1 in Newey (1991) and *3.*, to prove the required uniform convergence, it remains to show $G_n(\rho)$ to $\overline{G}_n(\rho)$ pointwisely. By proposition 3, Gallant (1987, p.509) and *1.*, $\log f_{t,0}(\rho) - E(\log f_{t,0}(\rho)), t = 1, 2, 3, \ldots$ is a mixingale of size $-\frac{2(r-2)}{r-4}$. Then by proposition 4, Gallant (1987, p.515), the pointwise convergence follows from (2). □

[3] For the definition of near epoch dependence of random functions, see Gallant (1987,p.497)

[4] For the definition of stochastic equicontinuity, see Newey (1991) or Andrews (1994)

[5] For example, see Newey and McFadden (1994)

Theorem 5. *For $\forall t \in \mathbb{Z}, \forall L \in \mathbb{Z}_+$, let $S_{t,t-L}(\rho) \equiv \frac{\mathrm{d}\log f_{t,t-L}(\rho)}{\mathrm{d}\rho}$, $I_{t,t-L}(\rho) \equiv E(S^2_{t,t-L}(\rho)) < \infty$. From the Regularity Conditions we immediately have $I_{t,t-L}(\rho) = E(-\frac{\mathrm{d}S_{t,t-L}(\rho)}{\mathrm{d}\rho})$. Assume the conditions in Theorem 2.1. hold and additionally, the following conditions also hold:*

1. *$\{\frac{1}{n}\sum_{t=1}^{n} E(I_{t,0}(\rho))\}_{n=1}^{\infty}$ is equicontinuous and $\{\frac{1}{n}\sum_{t=1}^{n}\frac{\mathrm{d}S_{t,0}(\rho)}{\mathrm{d}\rho}\}_{n=1}^{\infty}$ is stochastically equicontinuous on Λ;*
2. *For $\forall \rho \in \Lambda$, $\frac{\mathrm{d}S_{t,0}(\rho)}{\mathrm{d}\rho}$ is near epoch dependent of size $-\frac{2(r-2)}{r-4}$, $\sum_{t=1}^{\infty}\frac{\left\|\frac{\mathrm{d}S_{t,0}(\rho)}{\mathrm{d}\rho}\right\|_r}{t^2} < \infty$;*
3. *For $\forall \rho \in \Lambda$, $H(\rho) \equiv \lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n} I_{t,0}(\rho)$ exists and it is finite, continuous at $\rho = \rho_0$ and $H(\rho_0) > 0$, $\lim_{n\to\infty}\sup_{\rho\in\Lambda}\left|\frac{1}{n}\sum_{t=1}^{n} I_{t,0}(\rho) - H(\rho)\right| = 0$;*
4. *$\{\frac{1}{n}S_{t,0}(\rho) : t = 1, 2, 3, \ldots, n,\ n = 1, 2, 3, \ldots\}$ is a near epoch dependent array [6] of size $-\frac{1}{2}$.*
5. *Let $\sigma_n \equiv \frac{1}{n}\sqrt{\sum_{t=1}^{n} I_{t,0}(\rho_0)}$, $w_n(s) \equiv \frac{1}{\sqrt{\sum_{t=1}^{n} I_{t,0}(\rho_0)}}\sum_{t=1}^{[ns]} S_{t,0}(\rho_0)$, $\lim_{n\to\infty} Var(w_n(s)) = s$, $0 \le s \le 1$;*
6. *$\left\|\frac{1}{n}S_{t,0}(\rho_0)\right\|_r \le M < \infty$, where M is a constant, $t = 1, 2, \ldots, n$, $\quad n \in \mathbb{Z}_+$.*

Then $\sqrt{n}(\hat{\rho} - \rho_0) \xrightarrow{d} N(0, V_{MLE})$ where $V_{MLE} \equiv \frac{1}{H(\rho_0)}$.

Proof. A standard argument[7] for the asymptotic normality of an extremum estimator consists of two parts: the uniform convergence of second order conditions and the asymptotic normality of first order conditions. As in the proof of Theorem 4, combining *1.*, *2.* and *3.* we get $\sup_{\rho\in\Lambda}\left|\frac{1}{n}\sum_{t=1}^{n}\frac{\mathrm{d}S_{t,0}(\rho)}{\mathrm{d}\rho} - (-H(\rho))\right| \xrightarrow{p} 0$. By Theorem 2, Gallant(1987, p.519), *3.*, *4.*, *5.* and *6.* imply that $\frac{1}{\sqrt{n}}\sum_{t=1}^{n} S_{t,0}(\rho_0) \xrightarrow{d} N(0, H(\rho_0))$. □

[6] For the definition of a near epoch dependent array of random functions, see Gallant (1987, p.497)
[7] For example, see Newey and McFadden (1994)

3.2.2 Asymptotic Properties of the Markovian QMLE

Compared to the previous MLE, a distinctive feature in the Markovian QMLE case is that for any given $L \in \mathbb{Z}_+$, for $\forall \rho \in \Lambda$, $\{f_{t,t-L}(\rho)\}_{t=-\infty}^{\infty}$, $\{S_{t,t-L}\rho\}_{t=-\infty}^{\infty}$ and $\{\frac{\mathrm{d}S_{t,t-L}\rho}{\mathrm{d}\rho}\}_{t=-\infty}^{\infty}$ inherit both the stationarity and the strong mixing property of size $-\frac{4r}{r-4}$ for some $r > 4$ from $\{Y_t\}_{-\infty}^{\infty}$. This fact allows us to replace near epoch dependence conditions with suitable boundedness conditions of higher order moments.

Theorem 6. *Assume the Regularity Conditions, the Mixing Condition and the following hold:*

1. *For* $\forall \rho \in \Lambda$, $E\Big|\log f_{t,t-L}(\rho)\Big|^4 < \infty$;
2. $Q_n(\rho) \equiv \frac{1}{n}\sum_{t=L}^{n} \log f_{t,t-L}(\rho), n = L, L+1, \ldots$ *is stochastically equicontinuous.*

Then $\tilde{\rho}_n \xrightarrow{p} \rho_0$.

Proof. By the Regularity Conditions, it suffices to show the uniform convergnece of $Q_n(\rho)$ to $E(\log f_{t,t-L}(\rho))$. By Theorem 2.1, Newey(1991), based on *2.* it is reduced to show the pointwise convergence. First note that $\sum_{s=1}^{\infty}(\alpha(s))^{\frac{1}{2}} < \infty$ where $\alpha(s)$ is the mixing coefficient of $\mathcal{F}_{-\infty}^{t}$ and by assumption it is of size $-\frac{4r}{r-4}, r > 4$. By *1.* and Theorem 17.2.2, Ibragimov and Linnik(1971), for $\forall \rho \in \Lambda$, $\Big|Cov(\log f_{t,t-L}(\rho),\ \log f_{s,s-L}(\rho))\Big| < C(\alpha(|t - s|))^{\frac{1}{2}}$ where $C < \infty$ is a constant. Then $E\Big[\frac{1}{n-L+1}\sum_{t=L}^{n} \log f_{t,t-L}(\rho) - E(\log f_{t,t-L}(\rho))\Big]^2 < \frac{1}{(n-L+1)^2}\Big[(n-L+1)Var(\log f_{t,t-L}(\rho)) + 2C\sum_{s=1}^{\infty}(\alpha(s))^{\frac{1}{2}}\Big] \to 0$ when $n \to \infty$. The desired pointwise convergence then follows by Markov inequality. □

Theorem 7. *Assume the conditions in Theorem 2.3. hold and additionally, the following also hold:*

1. *For* $\forall \rho \in \Lambda$, $E\Big((\frac{\mathrm{d}S_{t,t-L}(\rho)}{\mathrm{d}\rho})^4\Big) < \infty$;
2. $J_n(\rho) \equiv \frac{1}{n}\sum_{t=L}^{n} \frac{\mathrm{d}S_{t,t-L}(\rho)}{\mathrm{d}\rho}, n = L, L+1, \ldots,$ *is stochastically equicontinuous;*
3. $E(S_{t,t-L}(\rho)^2)$ *is continuous on* Λ *and* $E(S_{t,t-L}(\rho_0)^2) > 0$.

Then $\sqrt{n}(\tilde{\rho}_n - \rho_0) \xrightarrow{d} N(0, V_{QMLE})$ *where*

$$V_{QMLE} \equiv \frac{1}{\Big[E(S_{t,t-L}(\rho_0)^2)\Big]^2}\Big[E(S_{t,t-L}(\rho_0)^2) + 2\sum_{\tau=1}^{\infty} E\Big(S_{t,t-L}(\rho_0)S_{t-\tau,t-\tau-L}(\rho_0)\Big)\Big]$$

Proof. As in the proof of Theorem 5, it suffices to show the uniform convergence of second order conditions and the asymptotic normality of first order conditions. For the former purpose, the argument in Theorem 6 still applies. The pointwise convergence of $\frac{\mathrm{d}S_{t,t-L}(\rho)}{\mathrm{d}\rho}$ to $-E(S_{t,t-L}(\rho)^2)$ follows by *1.*, the Mixing condition and Markov inequality, just as in the proof of Theorem 6. Then by *2.* and *3.* we obtain the uniform convergence of $\frac{\mathrm{d}S_{t,t-L}(\rho)}{\mathrm{d}\rho}$ to $-E(S_{t,t-L}(\rho)^2)$. As the for the asymptotic normality of score functions, first note that $S_{t,t-L}(\rho_0), t = L, L+1, \dots$ is stationary, strong mixing of size $\frac{-4r}{r-4}$. Also (1) implies $E(S_{t,t-L}(\rho)^4) < \infty$. Then by Theorem 1.7, Ibragimov (1962), we have $\frac{1}{\sqrt{n}}\sum_{t=L}^{n} S_{t,t-L}(\rho_0) \xrightarrow{d} N(0, A)$ where $A \equiv E(S_{t,t-L}(\rho_0)^2) + 2\sum_{\tau=1}^{\infty} E\Big(S_{t,t-L}(\rho_0)S_{t-\tau,t-\tau-L}(\rho_0)\Big)$.

□

3.3 Equivalency Between the MLE and the Markovian QMLE

The previous section shows that the asymptotic distributions of the MLE and the QMLE are obviously different. This is expected, since the likelihood function used in QMLE is just “quasi”-likelihood, which is associated with loss of information as the history on which the likelihood function is conditional is truncated at the L-th lag. For a non-Markov process, this truncation will certainly cause loss of information.

In light of this interpretation, a natural question then arises: can the QMLE approximate the MLE when this truncation of history gradually disappears? In other words, will the asymptotic variances of the two estimators become equal when $L \to \infty$? This is what this section is focused on and it turns out that they do become equal when $L \to \infty$

under some mild boundedness conditions.

The general idea is to break down this problem into two parts:

- (a) $0 < \lim_{L\to\infty} E(S_{t,t-L}(\rho_0)^2) = \frac{1}{V_{MLE}} = \lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} E(S_{t,0}(\rho_0)^2) < \infty$;
- (b) $\lim_{L\to\infty} \sum_{\tau=1}^{\infty} E\Big(S_{t,t-L}(\rho_0) S_{t-\tau,t-\tau-L}(\rho_0)\Big) = 0$.

For the part $\lim_{L\to\infty} E(S_{t,t-L}(\rho_0)^2) = \frac{1}{V_{MLE}} = \lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} E(S_{t,0}(\rho_0)^2) < \infty$ in (a), by stationarity an obvious equivalent condition is that $\lim_{L\to\infty} E(S_{t,t-L}(\rho_0)^2)$ exists and finite. For the "$0 <$" part, the following proposition provides a simple necessary condition.

Denote the unconditional score function by $M_t \equiv \frac{\mathring{d}p_{t,0}(\rho_0)}{\mathrm{d}\rho}$. By definition, $M_t - M_{t-1} = S_{t,0}(\rho_0)$. It is a well-known fact that if $\forall t \in \mathbb{Z}_+, E|M_t| < \infty$, then $M_{t,0}, t = 1, 2, \dots$ is a martingale with respect to $\mathcal{F}_0^t \equiv \sigma(Y_0, Y_1, \dots, Y_t)$.

Proposition 5. *If* $\lim_{L\to\infty} E(S_{t,t-L}(\rho_0)^2)$ *exists, finite and strictly positive, and* $\forall t \in \mathbb{Z}_+, E|M_t| < \infty$, *then* $\lim_{t\to\infty} E(M_t^2) = \infty$.

Proof. By stationarity and the martingale property of $M_{t,0}$, $\lim_{L\to\infty} E(S_{t,t-L}(\rho_0)^2) = \lim_{t\to\infty} E(S_{t,0}(\rho_0)^2)$ $= \lim_{t\to\infty} \Big[E(M_t^2) - E(M_{t-1}^2)\Big]$. If $\lim_{t\to\infty} E(M_t^2) = M < \infty$, then $\lim_{t\to\infty} \Big[E(M_t^2) - E(M_{t-1}^2)\Big] = 0$, contradicting that $\lim_{L\to\infty} E(S_{t,t-L}(\rho_0)^2)$ is strictly positive. □

$\lim_{t\to\infty} E(M_t^2) = \infty$ is actually a common regularity assumption in the literature of inference on stochastic processes, e.g., see Section 6.2 in Hall and Heyde (1980).

For (b), things get more complicated. Let $V_{L,\tau} \equiv E\Big(S_{t,t-L}(\rho_0) S_{t-\tau,t-\tau-L}(\rho_0)\Big)$ and $S_{t,t-L} \equiv S_{t,t-L}(\rho_0)$ for the simplicity of notations. First note that if conditions of the dominated convergence theorem are satisfied, then proving (b) holds is reduced to proving $\lim_{L\to\infty} V_{L,\tau} = 0$ for any given τ. Unfortunately, it seems hard to me to formulate a primitive condition for the existence of an integrable dominating function, so its existence is assumed in this paper: [8]:

Assumption 1 $\sum_{\tau=1}^{\infty} \sup_{L\geq 1} V_{L,\tau} < \infty$

[8]In fact, a weaker condition is also sufficient: if there exists a sequence of integrable functions $g_L(\tau)$ such that $0 \le V_{L,\tau} \le g_L(\tau)$and $\liminf_{L\to\infty} \sum_{\tau=1}^{\infty} g_L(\tau) = \sum_{\tau=1}^{\infty} \liminf_{L\to\infty} g_L(\tau) < \infty, L = 1, 2, \dots, \tau = 1, 2, \dots$

With Assumption 1 at hand, what remains to do is proving for any given τ, $\lim_{L\to\infty} V_{L,\tau} = 0$. For $s \leq t$, denote $\mathcal{F}_s^t \equiv \sigma(Y_t, Y_{t-1}, \dots, Y_s)$. To get an overview of the strategy, first observe that suppose for some $L \in \mathbb{Z}_+$, $(Y_t, Y_{t-1}, \dots, Y_{t-L})$ and $(Y_{t-\tau}, Y_{t-\tau-1}, \dots, Y_{t-\tau-L})$ are independent conditional on $\mathcal{F}_{t-L}^{t-1}$, then by law of iterated expectation and the martingale difference property of conditional score functions, $V_{L,\tau} = E\Big(E(S_{t,t-L}S_{t-\tau,t-\tau-L}|\mathcal{F}_{t-L}^{t-1})\Big) = E\Big(E(S_{t,t-L}|\mathcal{F}_{t-L}^{t-1})E(S_{t-\tau,t-\tau-L}|\mathcal{F}_{t-L}^{t-1})\Big) = E\Big(0 \cdot E(S_{t-\tau,t-\tau-L}|\mathcal{F}_{t-L}^{t-1})\Big) = 0$. Following this observation, a natural conjecture is that:

If the conditional independence holds when $L = \infty$, then $\lim_{L\to\infty} V_{L,\tau} = 0$.

To make discussion rigorous, it is necessary to define what it means by "conditional independence when $L = \infty$".

Definition 8. Let $\mathcal{F}_n, \mathcal{G}_n, \mathcal{H}_n$ be sequences of σ-algebras, $n = 1, 2, \dots$ If when $n \to \infty$,

$$\sup_{F_n \in \mathcal{F}_n \vee \mathcal{G}_n, H_n \in \mathcal{H}_n \vee \mathcal{G}_n} E\Big|E(\boldsymbol{1}_{F_n}\boldsymbol{1}_{H_n}|\mathcal{G}_n) - E(\boldsymbol{1}_{F_n}|\mathcal{G}_n)E(\boldsymbol{1}_{H_n}|\mathcal{G}_n)\Big| \to 0$$

where $\mathcal{F}_n \vee \mathcal{G}_n$ and $\mathcal{H}_n \vee \mathcal{G}_n$ mean the σ-algebra generated by $\{\mathcal{F}_n, \mathcal{G}_n\}$ and $\{\mathcal{H}_n, \mathcal{G}_n\}$ respectively, then we call $\mathcal{F}_n$ and $\mathcal{H}_n$ are *asymptotically independent conditional on* $\mathcal{G}_n$.

The following theorem illustrates the potential use of this definition.

Theorem 9. *Suppose $\mathcal{F}, \mathcal{G}$ and $\mathcal{H}$ are three arbitrary σ-algebras , X is $\mathcal{F}\vee\mathcal{G}$-measurable and Y is $\mathcal{H}\vee\mathcal{G}$-measurable, $|X| \leq C_1$ a.s., $|Y| \leq C_2$ a.s.. Then*

$$E\Big|E(XY|\mathcal{G}) - E(X|\mathcal{G})E(Y|\mathcal{G})\Big| \leq 4C_1C_2 \sup_{F\in\mathcal{F}\vee\mathcal{G}, H\in\mathcal{H}\vee\mathcal{G}} E\Big|E(\mathbf{1}_F\mathbf{1}_H|\mathcal{G}) - E(\mathbf{1}_F|\mathcal{G})E(\mathbf{1}_H|\mathcal{G})\Big|$$

Proof. The proof basically follows that of Theorem A.5, Hall and Heyde(1980). Let $\eta \equiv sgn\Big\{E(Y|\mathcal{G}\vee\mathcal{F}) - E(Y|\mathcal{G})\Big\}$, then by definition η is $\mathcal{G}\vee\mathcal{F}$-measurable. By the law of iterated expectations, $E\Big|E(XY|\mathcal{G}) - E(X|\mathcal{G})E(Y|\mathcal{G})\Big| = E\Big|E\Big[X\Big(E(Y|\mathcal{G}\vee\mathcal{F}) - E(Y|\mathcal{G})\Big)\Big|\mathcal{G}\Big]\Big| \leq E\Big|X\Big(E(Y|\mathcal{G}\vee\mathcal{F}) - E(Y|\mathcal{G})\Big)\Big| \leq C_1E\Big|E(Y|\mathcal{G}\vee\mathcal{F}) - E(Y|\mathcal{G})\Big| = C_1E\Big(\eta E(Y|\mathcal{G}\vee\mathcal{F}) - \eta E(Y|\mathcal{G})\Big) = C_1E\Big[E\Big(E(\eta Y|\mathcal{G}\vee\mathcal{F}) - \eta E(Y|\mathcal{G})\Big|\mathcal{G}\Big)\Big] = C_1E\Big[E(\eta Y|\mathcal{G}) - E(\eta|\mathcal{G})E(Y|\mathcal{G})\Big] \leq C_1E\Big|E(\eta Y|\mathcal{G}) - E(\eta|\mathcal{G})E(Y|\mathcal{G})\Big|$.

Furthermore, let $\xi \equiv sgn\Big\{E(\eta|\mathcal{G} \vee \mathcal{H}) - E(\eta|\mathcal{G})\Big\}$, which is $\mathcal{G} \vee \mathcal{H}$-measurable. By a similar reduction we have $C_1 E\Big|E(\eta Y|\mathcal{G}) - E(\eta|\mathcal{G})E(Y|\mathcal{G})\Big| \le C_2 C_1 E\Big|E(\eta\xi|\mathcal{G}) - E(\eta|\mathcal{G})E(\xi|\mathcal{G})\Big|$.

Now set $A_1 \equiv \{\eta = 1\}$, $A_2 \equiv \{\eta = -1\}$, $B_1 \equiv \{\xi = 1\}$, $B_2 \equiv \{\xi = -1\}$. Then $E\Big|E(\eta\xi|\mathcal{G}) - E(\eta|\mathcal{G})E(\xi|\mathcal{G})\Big| = E\Big|E((\boldsymbol{1}_{A_1} - \boldsymbol{1}_{A_2})(\boldsymbol{1}_{B_1} - \boldsymbol{1}_{B_2})|\mathcal{G}) - E(\boldsymbol{1}_{A_1} - \boldsymbol{1}_{A_2}|\mathcal{G})E(\boldsymbol{1}_{B_1} - \boldsymbol{1}_{B_2}|\mathcal{G})\Big| = E\Big|E(\boldsymbol{1}_{A_1}\boldsymbol{1}_{B_1}|\mathcal{G}) - E(\boldsymbol{1}_{A_1}|\mathcal{G})E(\boldsymbol{1}_{B_1}|\mathcal{G}) + E(\boldsymbol{1}_{A_2}\boldsymbol{1}_{B_2}|\mathcal{G}) - E(\boldsymbol{1}_{A_2}|\mathcal{G})E(\boldsymbol{1}_{B_2}|\mathcal{G}) - [E(\boldsymbol{1}_{A_1}\boldsymbol{1}_{B_2}|\mathcal{G}) - E(\boldsymbol{1}_{A_1}|\mathcal{G})E(\boldsymbol{1}_{B_2}|\mathcal{G})] - [E(\boldsymbol{1}_{A_2}\boldsymbol{1}_{B_1}|\mathcal{G}) - E(\boldsymbol{1}_{A_2}|\mathcal{G})E(\boldsymbol{1}_{B_1}|\mathcal{G})]\Big|$
$\le \sup_{F \in \mathcal{F} \vee \mathcal{G}, H \in \mathcal{H} \vee \mathcal{G}} 4E\Big|E(\boldsymbol{1}_F \boldsymbol{1}_H|\mathcal{G}) - E(\boldsymbol{1}_F|\mathcal{G})E(\boldsymbol{1}_H|\mathcal{G})\Big|$. □

Let $\mathcal{F}_L \equiv \sigma(Y_t)$, $\mathcal{G}_L \equiv \sigma(Y_{t-1}, Y_{t-2}, \ldots, Y_{t-L})$, $\mathcal{H}_L \equiv \sigma(Y_{t-\tau}, Y_{t-\tau-1}, \ldots, Y_{t-\tau-L})$, where $t \in \mathbb{Z}$ and $\tau \ge 1$ are fixed. Obviously, $S_{t,t-L}$ is $\mathcal{F}_L \vee \mathcal{G}_L$-measurable and $S_{t-\tau,t-\tau-L}$ is $\mathcal{H}_L \vee \mathcal{G}_L$-measurable for $L = 1, 2, \ldots$. From the previous theorem, we immediately have the following corollary:

Corollary 10. Assume the following conditions hold:

1. $\mathcal{F}_L$ and $\mathcal{H}_L$ are asymptotically independent conditional on $\mathcal{G}_L$ in the sense of Definition 3.2.;

2. $|S_{t,t-L}| < C_1 < \infty$ a.s., $|S_{t-\tau,t-\tau-L}| < C_2 < \infty$ a.s. for all $L = 1, 2, \ldots$.

Then $\lim_{L\to\infty} V_{L,\tau} = 0$

Remark 9. *The condition that $\mathcal{F}_L$ and $\mathcal{H}_L$ are asymptotically independent conditional on $\mathcal{G}_L$ is much weaker than the usual Markov condition. In fact, it can be interpreted as an "asymptotic Markov condition": If $\{Y_t\}_{t=-\infty}^{\infty}$ is a Markov process, then $\forall t \in \mathbb{Z}$, $\sigma(Y_t)$ is independent of $\sigma(Y_{t-1}, Y_{t-2}, \ldots)$ conditional on $\sigma(Y_{t-1})$, which implies that for $\forall L \in \mathbb{Z}_+$, $\forall \tau \in \mathbb{Z}_+$, and $\forall t \in \mathbb{Z}$, $\sigma(Y_t, Y_{t-1}, \ldots, Y_{t-L})$ is independent of $\sigma(Y_{t-1}, Y_{t-2}, \ldots, Y_{t-\tau-L})$ conditional on $\sigma(Y_{t-1}, Y_{t-2}, \ldots, Y_{t-L})$. In other words, for $\forall L \in \mathbb{Z}_+$, $\forall \tau \in \mathbb{Z}_+$, and $\forall t \in \mathbb{Z}$,*

$$\sup_{F_L \in \mathcal{F}_L \vee \mathcal{G}_L, H_L \in \mathcal{H}_L \vee \mathcal{G}_L} E\Big|E(\mathbf{1}_{F_L}\mathbf{1}_{H_L}|\mathcal{G}_L) - E(\mathbf{1}_{F_L}|\mathcal{G}_L)E(\mathbf{1}_{H_L}|\mathcal{G}_L)\Big| = 0$$

Now a natural question is: under what conditions, the conditional asymptotic independence holds. It turns out in the case this paper is focused on, this condition holds under a compactness condition. Specifically, suppose we have an arbitrary process $\{Y_t\}_{t=-\infty}^{\infty}$, if $\mathcal{F}_L \equiv \sigma(Y_t)$, $\mathcal{G}_L \equiv \sigma(Y_{t-1}, Y_{t-2}, \ldots, Y_{t-L})$, $\mathcal{H}_L \equiv \sigma(Y_{t-\tau}, Y_{t-\tau-1}, \ldots, Y_{t-\tau-L})$, where $t \in \mathbb{Z}$ and $\tau \geq 1$ are fixed. Let $\mathcal{A}$ be the set of all measurable sets in the sample space Ω, $S \equiv \{\boldsymbol{1}_A, A \in \mathcal{A}\}$, and denote by $L^2(\Omega)$ the space of random variables which have finite second moment. If S is a compact subset of $L^2(\Omega)$, then $\mathcal{F}_L$ and $\mathcal{H}_L$ are asymptotically independent conditional $\mathcal{G}_L$ when $L \to \infty$. Note that this does not require $\{Y_t\}_{t=-\infty}^{\infty}$ to be stationary or strong mixing. Before proceeding to the proof, we need to state the following useful lemma:

Lemma 1. *Let $\mathcal{F}_n, \mathcal{G}_n, \mathcal{H}_n$ be sequences of σ-algebras, $n = 1, 2, \ldots$ If when $n \to \infty$,*

$$\sup_{H_n \in \mathcal{H}_n \vee \mathcal{G}_n} E\Big|E(\mathbf{1}_{H_n}|\mathcal{G}_n \vee \mathcal{F}_n) - E(\mathbf{1}_{H_n}|\mathcal{G}_n)\Big| \to 0$$

then $\mathcal{F}_n$ is asymptotically independent of $\mathcal{H}_n$ conditional on $\mathcal{G}_n$.

Proof. For $\forall F_n \in \mathcal{F}_n \vee \mathcal{G}_n$, $\forall H_n \in \mathcal{H}_n \vee \mathcal{G}_n$, $E\Big|E(\boldsymbol{1}_{F_n}\boldsymbol{1}_{H_n}|\mathcal{G}_n) - E(\boldsymbol{1}_{F_n}|\mathcal{G}_n)E(\boldsymbol{1}_{H_n}|\mathcal{G}_n)\Big| = E\Big|E\Big[E(\boldsymbol{1}_{H_n}|\mathcal{F}_n \vee \mathcal{G}_n)\boldsymbol{1}_{F_n}|\mathcal{G}_n\Big] - E(\boldsymbol{1}_{F_n}|\mathcal{G}_n)E(\boldsymbol{1}_{H_n}|\mathcal{G}_n)\Big| = E\Big|E\Big[\Big(E(\boldsymbol{1}_{H_n}|\mathcal{F}_n \vee \mathcal{G}_n) - E(\boldsymbol{1}_{H_n}|\mathcal{G}_n)\Big)\boldsymbol{1}_{F_n}|\mathcal{G}_n\Big]\Big|$
$\leq E\Big|\Big(E(\boldsymbol{1}_{H_n}|\mathcal{F}_n \vee \mathcal{G}_n) - E(\boldsymbol{1}_{H_n}|\mathcal{G}_n)\Big)\boldsymbol{1}_{F_n}\Big| \leq E\Big|E(\boldsymbol{1}_{H_n}|\mathcal{F}_n \vee \mathcal{G}_n) - E(\boldsymbol{1}_{H_n}|\mathcal{G}_n)\Big|$. □

Theorem 11. *If S is a compact subset of $L^2(\Omega)$, then $\mathcal{F}_L$ and $\mathcal{H}_L$ are asymptotically independent conditional on $\mathcal{G}_L$.*

Proof. In view of Lemma 1, it suffices to show that $\sup_{F_L \in \mathcal{F}_L \vee \mathcal{G}_L} E\Big|E(\boldsymbol{1}_{F_L}|\mathcal{G}_L \vee \mathcal{H}_L) - E(\boldsymbol{1}_{F_L}|\mathcal{G}_L)\Big| \to 0$ when $L \to \infty$. A further sufficient condition is that $\sup_{F \in S} E\Big|E(\boldsymbol{1}_F|\mathcal{G}_L \vee \mathcal{H}_L) - E(\boldsymbol{1}_F|\mathcal{G}_L)\Big|^2 \to 0$. The pointwise convergence follows immediately from the L^2-convergence property of martingales. Indeed, by Theorem 5.5, Varadhan(2001) for a given F, we have $E\Big|E(\boldsymbol{1}_F|\mathcal{G}_L) - E(\boldsymbol{1}_F|\mathcal{G}_\infty)\Big|^2 \to 0$. Note that $\mathcal{G}_L \vee \mathcal{H}_L = \mathcal{G}_{L+\tau}$, so $E\Big|E(\boldsymbol{1}_F|\mathcal{G}_L) - E(\boldsymbol{1}_F|\mathcal{G}_L \vee \mathcal{H}_L)\Big|^2 \to 0$.

To prove the uniformity of the above convergence, first note that for $\forall L \in \mathbb{Z}_+$, $E\Big|E(\cdot|\mathcal{G}_L) - E(\cdot|\mathcal{G}_\infty)\Big|^2 : L^2(\Omega) \to \mathbb{R}$ is a continuous mapping since $E(\cdot|\mathcal{G}_L) - E(\cdot|\mathcal{G}_\infty) : L^2(\Omega) \to L^2(\Omega)$ is a continuous operator and $E\Big|\cdot\Big|^2 : L^2(\Omega) \to \mathbb{R}$ is a continuous mapping. Furthermore, for $\forall F \in S$, $E\Big|E(\boldsymbol{1}_F|\mathcal{G}_L) - E(\boldsymbol{1}_F|\mathcal{G}_\infty)\Big|^2 \geq E\Big|E(\boldsymbol{1}_F|\mathcal{G}_{L+1}) - E(\boldsymbol{1}_F|\mathcal{G}_\infty)\Big|^2$. Then for $\forall \epsilon > 0$, let $O_L \equiv \{F \in S : E\Big|E(\boldsymbol{1}_F|\mathcal{G}_L) - E(\boldsymbol{1}_F|\mathcal{G}_\infty)\Big|^2 < \epsilon\}$ is an open set. The last inequality implies that $O_1 \subset O_2 \subset \cdots O_L \subset \cdots$, and the pointwise convergence implies that $\bigcup_{L=1}^\infty O_L$ is an open cover of S. By compactness of S, there exists some L_0 such that for $\forall L > L_0$, $O_L = S$. This proves the uniformity of the convergence. □

To summarize this section, we have

Proposition 6. *Combining Proposition 3.1., Corollary 3.4., and Theorem 3.7., if we have*

1. $\lim_{L\to\infty} E(S_{t,t-L}(\rho_0)^2)$ *exists, finite and strictly positive;*
2. *Assumption 1 holds;*
3. $|S_{t,t-L}| < C_1 < \infty$ *a.s.,* $|S_{t-\tau,t-\tau-L}| < C_2 < \infty$ *a.s. for all* $L = 1, 2, \ldots$, C_1 *and* C_2 *may depend on* τ*;*
4. S *is a compact subset of* $L^2(\Omega)$.

Then $\lim_{L\to\infty} V_{QMLE} = V_{MLE}$.

3.4 Conclusion

This chapter studies the asymptotic properties of the maximum likelihood estimator and the quasi-maximum likelihood estimator constructed from a Markov hypothesis when the data generating process is a stationary and strong mixing process parametrized by a scalar. The context is at a general level. In particular, no assumption about the functional form of likelihood functions are imposed. The cost of this generality is that some high level assumptions, like stochastic equicontinuity and near epoch dependence, are needed when deriving the asymptotics of the two estimators. Furthermore, this paper also examines the relation between the two asymptotic distributions and describes the conditions under which the two asymptotic variances tend to be equal when more and more lags are used in the Markovian QMLE.

Future research may explore primitive conditions for those listed in Proposition 6 and the relation between these conditions and what are used in theorems in Section 2. In fact, conditions in Proposition 6, especially *1.* and *4.*, are arguably interesting problems in their own right. It will be interesting if they turn out to be implied, or even partially implied, by the dependence conditions which are used to derive the asymptotics in Section 2.

Bibliography

[1] Andrews, Donald. W. K., 1994, Empirical Process Methods in Econometrics, Chapter 37, *Handbook of Econometrics*, Vol.4, North Holland

[2] Audrino,Francesco, Robert Huitema and Markus Ludwig, 2019, An Empirical Implementation of the Ross Recovery Theorem as a Prediction Device, *Journal of Financial Econometrics*, 1-22.

[3] Bakshi, Gurdip, Fousseni Chabi-Yo, and Xiaohui Gao Bakshi, 2018, A Recovery that We Can Trust? Deducing and Testing the Restrictions of the Recovery Theorem, *Review of Financial Studies* 31, 532555.

[4] Borovička, Jaroslav, Lars Peter Hansen and José A. Scheinkman, 2016, Misspecified Recovery, *Journal of Finance* 71, 2493-2544.

[5] Breeden, Douglas T., and Robert H. Litzenberger, 1978, Prices of State-Contingent Claims Implicit in Option Prices, *Journal of Business* 51, 621651.

[6] Britten–Jones, Mark and Anthony Neuberger, 2000, Option Prices, Implied Price Processes, and Stochastic Volatility, *Journal of Finance* 55, 839-866.

[7] Carr, Peter and Jiming Yu, 2012, Risk, Return and Ross Recovery, *Journal of Derivatives* 20, 38-59.

[8] Chernov, Mikhail and Eric Ghysels, 2000, A study towards a unified approach to the joint estimation of objective and risk neutral measures for the purpose of options valuation, *Journal of Financial Economics* 56, 407-458.

[9] Davidson, James, 1994, *Stochastic Limit Theory*, Oxford University Press, Oxford, UK

[10] Derman, Emanuel, and Iraj Kani, 1994, Riding on a Smile, *Risk* 7, 32-39.

[11] Derman, Emanuel, Iraj Kani and Neil Chriss, 1996, Implied Trinomial Trees of the Volatility Smile, Goldman Sachs Quantitative Strategies Research Notes.

[12] Derman, Emanuel, and Iraj Kani, 1998, Stochastic implied trees: Arbitrage pricing with stochastic term and strike structure of volatility, *International Journal of Theoretical and Applied Finance* 1, 722.

[13] Dillschneider, Yannick and Raimond Maurer, 2019, Functional Ross Recovery: Theoretical Results and Empirical Test, *Journal of Economic Dynamics & Control* 108, 103750

[14] Dubynskiy, Sergey, and Robert S. Goldstein, 2013, Recovering Drifts and Preference Parameters from Financial Derivatives, Working paper, University of Minnesota.

[15] Duffie, Darrell and Mark B. Garman, 1989, Intertemporal Arbitrage and the Markov Valuation of Securities, Working paper No.975, Stanford University.

[16] Dupire, Bruno, 1994, Pricing with a smile, *Risk* 7, 1820.

[17] Durrett, Rick, 2010, *Probability: Theory and Examples*(4th Edition), Cambridge University Press, Cambridge, UK.

[18] Gallant, A. Ronald, 1987, *Nonlinear Statistical Models*, Wiley, New York

[19] Gallant, A. Ronald and Jonathan Long, 1997, Estimating Stochastic Differential Equations Efficiently by Minimum Chi-Squared, *Biometrika*, 84, 125-141

[20] Gallant, A. Ronald and George Tauchen, 1996, Which moments to match?, *Econometric Theory* 12, 657-681.

[21] Gallant, A. Ronald and George Tauchen, 1998, Reprojecting partially observed systems with application to interest rate diffusions, *Journal of American Statistical Association* 93, 10-24.

[22] Hall, Peter and Chris C. Heyde, 1980, *Martingale Limit Theory and Its Application*, Academic Press, New York

[23] Hall, Peter and Qiwei Yao, 2003, Inference in ARCH and GARCH Models with Heavy-Tailed Errors, *Econometrica* 71, 283-317

[24] Hansen, Lars Peter and José Scheinkman, 2009, Long Term Risk: An Operator Approach, *Econometrica* 77, 177-234.

[25] Harrison, J. Michael and David M. Kreps, 1979, Martingales and Arbitrage in Multiperiod Securities Markets, *Journal of Economic Theory* 20, 381-408

[26] Harrison, J. Michael and Stanley R. Pliska, 1981, Martingales and Stochastic Integrals in the Theory of Continuous Trading, *Stochastic Processes and their Applications* 11, 215-260.

[27] Heyde, Chris C., 1997, *Quasi-Likelihood And Its Application: A General Approach to Optimal Parameter Estimation*, Springer, New York

[28] Ibragimov, I. A., 1962, Some Limit Theorems for Stationary Processes, *Theory of Probability and Its Applications* 7, 349-382

[29] Ibragimov, I. A. and Y. Linnik, 1971, *Independent and Stationary Sequences of Random Variables*, Wolters-Noordhoff

[30] Jackwerth, Jens Carsten, 1999, Option-Implied Risk-Neutral Distributions and Implied Binomial Trees: A Literature Review, *Journal of Derivatives* 7, 66-82.

[31] Jackwerth, Jens Carsten and Marco Menner, 2018, Does the Ross Recovery Theorem Work Empirically?, Working paper, University of Konstanz.

[32] Jensen, Christian Skov, David Lando and Lasse Heje Pedersen, 2019, Generalized Recovery, *Journal of Financial Economics* 133, 154-174.

[33] Kallenberg, Olav, 1997, *Foundations of Modern Probability* (Springer, New York, USA).

[34] Martin, Ian and Stephen A. Ross, 2019, Notes on the yield curve, *Journal of Financial Economics* 134, 689-702.

[35] Newey, Whitney K., 1991, Uniform Convergence in Probability and Stochastic Equicontinuity, *Econometrica*, 59, 1161-1167

[36] Newey, Whitney K., and Daniel McFadden, 1994, Large Sample Estimation and Hypothesis Testing, Chapter 36, *Handbook of Econometrics*, Vol.4, North Holland

[37] Qin, Likuan, and Vadim Linetsky, 2016, Positive Eigenfunctions of Markovian Pricing Operators: HansenScheinkman Factorization and Ross Recovery, *Operations Research* 64, 99117.

[38] Ross, Stephen A., 2015, The Recovery Theorem, *Journal of Finance* 70, 615-648.

[39] Rubinstein, Mark, 1994, Implied Binomial Trees, *Journal of Finance* 49, 771-818.

[40] Schneider, Paul and Fabio Trojani, 2019, (Almost) Model-Free Recovery, *Journal of Finance* 74, 323-370.

[41] Tran, Ngoc-Khanh, and Shixiang Xia, 2018, Specified Recovery, Working paper, Washington University in St. Louis.

[42] Tran, Ngoc-Khanh, 2019, The Functional Stochastic Discount Factor, forthcoming in *Quarterly Journal of Finance.*

[43] Varadhan, S. R. S., 2001, *Probability Theory*, American Mathematical Society, Providence, Rhode Island

[44] Walden, Johan, 2017, Recovery with Unbounded Diffusion Processes, *Review of Finance* 21, 14031444.

[45] Yao, Xiangkun, 2018, What Information Can Ross' Recovery Theory Recover?-Evidence from S&P 500 Index Option Prices, Working paper, University of Connecticut.

Vita

Shengbo Zhu

Pennsylvania State University
Department of Economics
303 Kern Bldg, University Park
State College, PA 16802, USA

Email: shbzhu@gmail.com
Phone: +1 (814) 409-8856
Website: sites.google.com/view/zhushengbo

Education

Ph.D. in Economics, Pennsylvania State University, USA, 2020 (expected)
M.A. in Economics, Renmin University of China, China, 2014
B.S. in Chemistry, Renmin University of China, China, 2013
B.S. in Mathematics, Renmin University of China, China, 2013
B.S. in Economics, Renmin University of China, China, 2012

Working Papers

1. Why Does the Recovery Theorem Not Work Empirically? An Intrinsic Explanation
2. General Implied Processes and Direct Implementation of the Fundamental Theorem of Asset Pricing
3. Markovian Quasi-Maximum Likelihood Estimation, *with A. Ronald Gallant*

Employment

Research Assistant for Prof. A. Ronald Gallant — *Summer 2017, 2018, 2019, 2020*

Course Instructor:
Introduction to Econometrics (online course, undergraduate) — June 2016

Teaching Assistant:
Introductory Microeconomics (undergraduate) — August 2014 - May 2016
Introduction to Econometrics (undergraduate) — August 2016 - May 2018
Environmental Economics (undergraduate) — August 2018 - May 2019

Presentations

Duke Financial Econometrics Lunch Group — *September 2019*
Econometrics Workshop at UNC Chapel Hill — *November 2019*