Abstract

Two Papers of Financial Engineering Relating to the Risks of the 2007-2008 Financial Crisis

Haowen Zhong

This dissertation studies two financial engineering and econometrics problems relating to two facets of the 2007-2008 financial crisis.

In the first part, we construct the Spatial Capital Asset Pricing Model and the Spatial Arbitrage Pricing Theory to characterize the risk premiums of futures contracts on real estate assets. We also provide rigorous econometric analysis of the new models. Empirical study shows there exists significant spatial interaction among the S&P/Case-Shiller Home Price Index futures returns.

In the second part, we perform empirical studies on the jump risk in the equity market. We propose a simple affine jump-diffusion model for equity returns, which seems to outperform existing ones (including models with Levy jumps) during the financial crisis and is at least as good during normal times, if model complexity is taken into account. In comparing the models, we made two empirical findings: (i) jump intensity seems to increase significantly during the financial crisis, while on average there appears to be little change of jump sizes; (ii) finite number of large jumps in returns for any finite time horizon seem to fit the data well both before and after the crisis.
Contents

List of Figures iii

List of Tables vi

Acknowledgements viii

Dedication x

1 Overview 1

2 Location, Location, Location: Econometric Analysis of Asset Pricing with Spatial Interaction 4

2.1 Introduction ............................................. 4

2.2 Preliminary ............................................. 8

2.3 The Spatial Capital Asset Pricing Model ......................... 10

2.4 The Spatial Arbitrage Pricing Theory .......................... 14

2.4.1 Asymptotic Arbitrage .................................... 15
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.3</td>
<td>QQ Plots of Simulation Study</td>
<td>60</td>
</tr>
<tr>
<td>3.4</td>
<td>S&amp;P 500: Normalized Model Residues vs Standard Normal</td>
<td>64</td>
</tr>
<tr>
<td>3.5</td>
<td>S&amp;P 500: QQ plots Before and During the Crisis</td>
<td>65</td>
</tr>
<tr>
<td>3.6</td>
<td>S&amp;P 500: Jumps</td>
<td>68</td>
</tr>
<tr>
<td>3.7</td>
<td>Nasdaq 100: Jumps</td>
<td>68</td>
</tr>
<tr>
<td>3.8</td>
<td>Nasdaq 100: Normalized Model Residues vs Standard Normal</td>
<td>71</td>
</tr>
<tr>
<td>3.9</td>
<td>Nasdaq 100: QQ plots Before and During the Crisis</td>
<td>72</td>
</tr>
</tbody>
</table>


## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>The mean and standard deviation of $\hat{\rho}$. The asymptotic standard deviations are estimated from the sample average of Hessian matrix (see (2.48)).</td>
<td>32</td>
</tr>
<tr>
<td>2.2</td>
<td>Simulation study of the sample adjusted $R^2$. We use the same model specification as that for Table 2.1, except that $\rho_0$ is fixed at 0.5 and two values of $\sigma_0^2$ (0.01 and 0.5) are used, respectively, for the two data sets. For each data set, the MLE of parameters is estimated for the model (2.29) under the constraint $\tilde{\alpha} = 0$ and then the sample adjusted $R^2$ for each element of $\tilde{r}$ is calculated and compared to its theoretical counterpart. It appears that the sample adjusted $R^2$ and its theoretical counterpart align well.</td>
<td>35</td>
</tr>
<tr>
<td>2.3</td>
<td>Robustness check of the empirical results.</td>
<td>41</td>
</tr>
<tr>
<td>3.1</td>
<td>The (right) quantiles for the Laplace and normalized t densities.</td>
<td>50</td>
</tr>
<tr>
<td>3.2</td>
<td>Conditional Posteriors, prior means and standard deviations for SV-DEJ.</td>
<td>54</td>
</tr>
<tr>
<td>3.3</td>
<td>Simulation study of Bayesian MCMC.</td>
<td>55</td>
</tr>
<tr>
<td>3.4</td>
<td>Deviance Information Criteria (DIC) on Model Fitting of Simulated Data.</td>
<td>58</td>
</tr>
<tr>
<td>3.5</td>
<td>Summary statistics of the data sets.</td>
<td>62</td>
</tr>
<tr>
<td>3.6</td>
<td>S&amp;P 500: Model Parameters Estimates</td>
<td>63</td>
</tr>
</tbody>
</table>
3.7  S&P 500: Kolmogorov-Smirnov Tests and Deviance Information Criteria  67
3.8  S&P 500: Change of Jump-Related Parameters  67
3.9  Nasdaq 100: Model Parameter Estimates  69
3.10 Nasdaq 100: Kolmogorov-Smirnov Tests and Deviance Information Criterion  70
3.11 Nasdaq 100: Change of Jump-Related Parameters  70
Acknowledgements

First and foremost, I would like to express my heartfelt gratitude to my Ph.D. advisor, Professor Steven Kou, whose insights and support have profoundly shaped both this dissertation and my long-term career plan. This dissertation has greatly benefitted from his vision in financial engineering, his taste of choosing interesting problems, his intuition in taking the suitable approach, his constant enlightenment, kind encouragement and adherence to high standards. Moreover, the countless hours of discussion with him that have helped to generate this work have also guided me to better understand the field and my position in research. He is the best advisor one can ever hope for.

I am grateful to Professors Ward Whitt, Paul Glasserman, Xuedong He and Tim Leung, for serving on my thesis committee and for providing invaluable suggestions at different stages of my Ph.D. study. I am indebted to my other collaborators, Professors Cindy Yu, Xianhua Peng and Ning Cai, for a variety of ideas, delightful discussions and assistance. I am also thankful to all the faculty members at Columbia from whom I have taken courses and learned so much.

My thanks are due to all my friends and fellow Ph.D. students in IEOR department for their companionship. It is as if yesterday that I first came to this program, and I can still remember how I first met every single one of them. It is impossible to enumerate the tears and laughters we share, all of which have become treasures in my memory. I also thank all the staff members of this department for their support and help.

Last but not least, I would like to thank my parents from the bottom of my heart, for their...
unfailing support and understanding throughout the years. I would also very much like to share my joy with my lovely wife Yixian Li, for her unconditional love and unabated trust during difficult times.
To Hong, Yixian and Alexa:

the so far three most important females in my life
Chapter 1

Overview

The 2007-2008 financial crisis has profound impact on the global financial market as well as on various sectors of the real economy. In order to prevent a similar disaster, it is now consensus among regulators, academics and practitioners that improved understanding is needed for the different kinds of risks in the financial system. In this spirit, this dissertation attempts to obtain better understanding of two different kinds of risks related to the recent financial crisis: (1) housing risk in residential home market, and (2) the jump risk in the equity market. We study the two kinds of risks by constructing new econometric models and analysing new data emerged from the Crisis.

The plummet of US residential home prices is considered one of the major causes of the recent financial crisis. To model real estate prices, spatial interaction is well-known to be important since housing prices are significantly affected by neighborhood prices. In the first part of the dissertation, we investigate how spatial interactions affect the housing risk
in the US residential home market. We also propose a Spatial Capital Asset Pricing model (S-CAPM) and a Spatial Arbitrage Pricing Theory (S-APT) that extend the classical asset pricing models and incorporate spatial interaction among asset returns. Furthermore, we give rigorous econometric analysis of the models: deriving conditions for the identifiability of the models, investigating asymptotic properties of the estimators and studying test statistics needed to implement the models. Finally, an empirical study of the futures contracts on the S&P/Case-Shiller Home Price Indices shows that spatial interaction is statistically significant.

The recent financial crisis has translated into an overly volatile equity market. In the second part of the dissertation, we study jump risk in the equity market. We attempt to answer two questions: (i) How did jumps in equity returns change during the financial crisis; in particular, were there significant changes in jump rates or in jump sizes, or both? (ii) Were there finite number of large jumps (e.g. those in affine jump-diffusion models) or infinite number of small jumps (e.g. those in Lévy type models) in equity returns before and during the crisis? To answer these questions, we first find that a simple affine jump-diffusion model fits both S&P 500 and Nasdaq 100 daily return data well; the model outperforms existing ones (in particular models with Lévy jumps) during the crisis, and is at least as good before the crisis. Based on the model and the data sets, we conclude that: (i) Both positive and negative jump rates increased significantly during the financial crisis, while, somewhat surprisingly, there is little evidence that jump sizes have changed
on average before and after the crisis. (ii) The empirical evidence favors finite number of large jumps in equity returns over a finite-time horizon.
Chapter 2

Location, Location, Location:

Econometric Analysis of Asset Pricing with Spatial Interaction

2.1 Introduction

Spatial interaction is important in modeling real estate assets, as housing prices are significantly affected by neighborhood prices. Spatial interaction has been well studied in the spatial econometrics literature (see, e.g., Anselin, 1988; Cressie, 1993); housing prices in equilibrium are studied in Ortalo-Magné and Prat (2010) who propose a spatial asset pric-
ing model that takes into account endogenous location choices of households. Instead of studying prices of houses and apartments that are illiquid and difficult to be sold short, the present chapter concerns the risk and return of real estate securities that are liquid and can be easily shorted, such as futures contracts on the S&P/Case-Shiller Home Price Indices (CSI Indices) constructed by Case and Shiller (1987). More precisely, we attempt to connect spatial econometrics, which emphasizes the statistical modeling of spatial interaction, with classical asset pricing models including the capital asset pricing model (CAPM) and the arbitrage pricing theory (APT), which characterize risk-return relationship of financial assets that can be freely traded, especially be sold short.

As the present chapter focuses on real estate securities, it is in spirit closest to Deng, Quigley, and Van Order (2000), who study the extent to which option pricing theory can explain mortgage default and prepayment behavior, as well as the source of risk due to investor heterogeneity. It is important to study the risk-return relationship of real estate securities such as the CSI Indices futures because they are useful instruments for risk management and for hedging in residential housing markets (Shiller, 1993), similar to the function that futures contracts fulfill in other financial markets. The interaction of the risk-return relationships of real estate assets and financial assets is studied in Fan, Huszár, and Zhang (2013).

In the spirit of Pinkse and Slade (2010), the purpose of this chapter is not to extend the techniques of spatial econometrics without economic rationale. Indeed, the spatial econometric model proposed in this chapter are based on a rigorous theory for the risk-return
relationship of real estate securities. Following Manski (1993, 2007), we study both the identification and the statistical inference problems for the proposed spatial econometric model. Manski (2007, pp. 3–4) points out that “it is useful to separate the inferential problem into statistical and identification components” and that “the study of identification logically comes first” because “negative identification findings imply that statistical inference is fruitless: it makes no sense to try to use a sample of finite size to infer something that could not be learned even if a sample of infinite size were available.” See Section 2.5.1.

There are three major differences between the model proposed in this chapter and existing models in the literature: (i) The fundamental difference lies in that, as is shown in this chapter, the economic rationale of equilibrium pricing and no arbitrage pricing imposes certain constraints on the parameters in the econometric models; these constraints are the manifestation of both the effect of spatial interaction and the economic rationale of asset pricing. In contrast, the parameters in existing spatial econometric models are not subject to such constraints. (ii) The maximum likelihood estimator (MLE) of our model has a square root convergence rate, which contrasts to the slow rate of convergence of the MLE of spatial autogressive (SAR) model under certain spatial scenarios, as investigated in detail by Lee (2004). The MLE of our model has a square root convergence rate because the asymptotic properties of the MLE are obtained with the number of cross-sectional units $n$ being fixed and with the time length $T$ going to infinity, as opposed to the case of SAR models, in which $n$ grows to infinity. A detailed comparison between our model and the SAR model is provided in Section 2.4.4. (iii) The model proposed in this chapter incor-
porates three features: (a) a spatial lag in the dependent variables, (b) individual-specific fixed effects, and (c) heterogeneity of factor loadings on common factors; however, existing models have as yet incorporated only some but not all of these features. For example, Lee and Yu (2010a) investigate the asymptotic properties of the quasi-maximum likelihood estimators (QMLES) for spatial panel data models that incorporate the features (a) and (b) but not (c); Holly, Pesaran, and Yamagata (2010) and Pesaran and Tosetti (2011) consider panel data models that incorporate spatially correlated cross-section errors and the features (b) and (c), but not (a); see Anselin, Le Gallo, and Jayet (2008) and Lee and Yu (2010b) for more comprehensive discussion of spatial panel data models and the asymptotic properties of MLE and QMLE for these models.

The contribution of this chapter is threefold: (i) Theoretically, we extend the classical asset pricing theories of CAPM and APT by proposing a spatial CAPM (S-CAPM) and a spatial APT (S-APT) that incorporate spatial interaction. The S-CAPM and S-APT characterize how spatial interaction affects asset returns by assuming, respectively, that investors hold mean variance efficient portfolios and that there is no asymptotic arbitrage. (ii) From the viewpoint of econometric methodology, we derive the necessary and sufficient condition for the identifiability of parameters in the proposed models; furthermore, we derive the asymptotic properties of the MLE and the likelihood ratio test statistic needed to implement the models. Interestingly, the MLE of the proposed model has a square root rate of convergence, whereas the MLEs of some other spatial econometrical models may not have such convergence rates. (iii) We conduct an empirical study on the futures contracts
written on the CSI Indices; the study shows that the proposed S-APT theory is not rejected
and that the spatial interaction parameter in the model for the CSI Indices futures returns
corresponding to the ten U.S. cities is statistically significant.

The remainder of the chapter is organized as follows. In Section 2.2, a linear model
with spatial interaction is introduced. The S-CAPM and S-APT for ordinary assets and
futures contracts are derived in Sections 2.3 and 2.4, respectively. Section 2.5 studies
the identifiability condition and the asymptotic properties of the MLE, as well as the test
statistic for implementing the S-APT. An empirical study of the CSI Indices futures using
the S-APT is provided in Section 2.6. Section 2.7 concludes.

2.2 Preliminary

Consider a one-period economy with \( n \) risky assets in the market whose returns are
governed by the following linear model:

\[
  r_i = \rho \sum_{j=1}^{n} w_{ij} r_j + \alpha_i + \epsilon_i, \quad i = 1, \ldots, n,
\]

(2.1)

where \( r_i \) is the return of asset \( i \), \( \alpha_i \) is a constant, and \( \epsilon_i \) is the residual noise related to asset \( i \).

For \( i \neq j \), \( w_{ij} \) specifies the influence of the return of asset \( j \) on that of asset \( i \) due to spatial
interaction; and \( w_{ii} = 0 \). The degree of spatial interaction is represented by the parameter
\( \rho \). Let \( \tilde{r} := (r_1, \ldots, r_n)' \), \( W := (w_{ij}) \), \( \alpha := (\alpha_1, \ldots, \alpha_n)' \), and \( \tilde{\epsilon} := (\epsilon_1, \ldots, \epsilon_n)' \). Then, the
above model can be represented as
\[
\tilde{r} = \rho W \tilde{r} + \alpha + \tilde{\epsilon}, \quad E[\tilde{\epsilon}] = 0, \quad E[\tilde{\epsilon} \tilde{\epsilon}'] = V.
\] (2.2)

Following the convention in spatial econometrics, we assume that the spatial weight matrix \(W\) is exogenously given. \(W\) is typically defined using quantities related to the location of assets, such as distance, contiguity, and relative length of common borders. For instance, \(W\) can be specified as \(w_{ii} = 0\) and \(w_{ij} = d_{ij}^{-1}\) for \(i \neq j\), where \(d_{ij}\) is the distance between asset \(i\) and asset \(j\). If other asset returns do not have spatial influence on \(r_i\), then the \(i\)th row of \(W\) can simply be set to zero.

Henceforth, we assume that \(\rho^{-1}\) is not an eigenvalue of \(W\). Then, \(I_n - \rho W\) is invertible\(^1\) and (2.2) can be rewritten as
\[
\tilde{r} = (I_n - \rho W)^{-1} \alpha + (I_n - \rho W)^{-1} \tilde{\epsilon},
\] (2.3)

where \(I_n\) is the \(n \times n\) identity matrix. The mean and covariance matrix of \(\tilde{r}\) are thus given by
\[
\mu = E[\tilde{r}] = (I_n - \rho W)^{-1} \alpha, \quad \Sigma = Cov(\tilde{r}) = (I_n - \rho W)^{-1} V (I_n - \rho W')^{-1}.
\] (2.4)

\(^1\)Let \(\det(\cdot)\) denote matrix determinant and \(\omega_1, \ldots, \omega_n\) be the eigenvalues of \(W\). Then, \(\det(I_n - \rho W) = \prod_{j=1}^n (1 - \rho \omega_j) \neq 0\) if and only if \(\rho^{-1}\) is not an eigenvalue of \(W\).
2.3 The Spatial Capital Asset Pricing Model

In this section we develop a spatial capital asset pricing model (S-CAPM) that general-
izes the CAPM by incorporating spatial interaction. In our study, it is important to consider
futures contracts as stand-alone securities rather than as derivatives of the underlying in-
struments because the instruments underlying futures contracts in the real estate markets
may not be tradable. For example, the CSI Indices futures are traded at Chicago Mercantile
Exchange but the underlying CSI Indices cannot be traded directly.

Therefore, we develop the S-CAPM for both ordinary assets and futures contracts.
More specifically, suppose in the market there are \( n_1 \) ordinary risky assets with returns
\((r_1, \ldots, r_{n_1})\), a risk-free asset with return \( r \), and \( n_2 \) futures contracts. The return of a fu-
tures contract cannot be defined in the same way as that of an ordinary asset because the
initial value of a futures contract is zero. Hence, we follow the convention in the literature
(see, e.g., De Roon, Nijman, and Veld, 2000) and define

\[
r_{n_1+i} := \frac{F_{i,1} - F_{i,0}}{F_{i,0}}
\]

as the “nominal return” of the \( i \)th futures contract, where \( F_{i,0} \) and \( F_{i,1} \) are the futures prices
of the \( i \)th futures contract at time 0 and time 1 (the beginning and end of the trading pe-
riod), respectively, and \( i = 1, \ldots, n_2 \). Let \( n = n_1 + n_2 \) and assume that the \( n \) returns
\( \tilde{\mathbf{r}} = (r_1, \ldots, r_{n_1}, r_{n_1+1}, \ldots, r_n)' \) satisfy the model (2.2). Then, the mean \( \mu \) and covariance
matrix \( \Sigma \) of \( \tilde{\mathbf{r}} \) are given by (2.4).
Now consider the mean-variance problem faced by an investor who can invest in the \( n_1 \) ordinary assets and \( n_2 \) futures contracts. Because the investor’s portfolio includes both ordinary assets and futures contracts, the return of the portfolio has to be calculated more carefully than if there were no futures contracts in the portfolio. Then, the mean-variance analysis can be carried out; see Appendix A.1. Because both \( \mu \) and \( \Sigma \) are functions of \( \rho \) and \( W \), the optimal portfolio weights obtained by the mean-variance analysis and the efficient frontiers are affected by spatial interaction. For example, Figure 2.1 shows the efficient frontiers for different values of \( \rho \) with all of the other parameters in the model (2.2) fixed for a portfolio of ten assets. It is clear that the efficient frontiers are significantly affected by \( \rho \).

![Efficient frontiers for different values of \( \rho \).](image)

**Figure 2.1:** Efficient frontiers for \( \rho = 0, 0.2, 0.4, 0.6, \) and 0.8, respectively, when there is no risk-free asset. \( W, \alpha, \) and \( V \) are specified in Appendix A.1. The efficient frontiers are significantly affected by \( \rho \).
Based on the mean-variance analysis, we derive the following S-CAPM, which characterizes how spatial interaction affects expected asset return under market equilibrium.

**Theorem 2.3.1** (S-CAPM for Both Ordinary Assets and Futures) Suppose that there exists a risk-free return \( r \) and that the \( n = n_1 + n_2 \) risky returns satisfy the model (2.2), of which the first \( n_1 \) are returns of ordinary assets and the others are returns of futures contracts.

Suppose \( n_1 > 0 \). \(^2\) Let \( r_M \) be the return of market portfolio. If each investor holds a mean-variance efficient portfolio, then, in equilibrium, \( r_M \) is mean-variance efficient and every investor holds the market portfolio and the risk-free asset. Furthermore,

(i) for the ordinary assets,

\[
E[r_i] - r = \frac{\text{Cov}(r_i, r_M)}{\text{Var}(r_M)} (E[r_M] - r) = \frac{\phi'_M \Sigma \eta_i}{\phi'_M \Sigma \phi_M} (E[r_M] - r), \quad i = 1, \ldots, n_1; \tag{2.6}
\]

(ii) for the futures contracts,

\[
E[F_{i,1}] - F_{i,0} = \frac{\text{Cov}(F_{i,1}, r_M)}{\text{Var}(r_M)} (E[r_M] - r) = F_{i,0} \frac{\phi'_M \Sigma \eta_{n_1+i}}{\phi'_M \Sigma \phi_M} (E[r_M] - r), \quad i = 1, \ldots, n_2, \tag{2.7}
\]

where \( \Sigma \) is the covariance matrix of \( \tilde{r} \); \( \phi_M \) is the portfolio weights of the market portfolio; and \( \eta_i \) is the \( n \)-dimensional vector with the \( i \)th element being 1 and all other elements being 0.

---

\(^2\)Since the aggregate position of all market participants in a futures contract is zero, \( n_1 \) needs to be positive in order to ensure that the return of the market portfolio is well defined.
Define
\[ 1_{n_1,n_2} := (1, \ldots, 1, 0, \ldots, 0)' \quad (2.8) \]
then \( \tilde{r} - r_{1_{n_1,n_2}} \) is the excess asset return\(^3\) and the S-CAPM equations (2.6) and (2.7) are equivalent to a single equation
\[
E[\tilde{r}] - r_{1_{n_1,n_2}} = \frac{Cov(\tilde{r}, r_M)}{Var(r_M)}(E[r_M] - r). \quad (2.9)
\]

**Proof.** See Appendix A.2.1.

The S-CAPM generalizes not only the CAPM for ordinary assets but also the CAPM for futures presented in Black (1976) and in Duffie (1989, Chapter 4) by incorporating spatial interaction. The S-CAPM can also be extended to the case in which there is no risk-free asset; see Appendix A.2.2.

It follows from the S-CAPM equations (2.6) and (2.7) that the degree of spatial interaction represented by the parameter \( \rho \) affects asset risk premiums in equilibrium because \( \Sigma \) is a function of \( W \) and \( \rho \) (see (2.4)).

The S-CAPM implies a constraint on the intercept of spatial econometric models for asset returns. Consider the following spatial econometric model, in which the excess returns \( \tilde{r} - r_{1_{n_1,n_2}} \) are regressed with a spatial interaction term on the excess return of the

\(^3\) \( \tilde{r} - r_{1_{n_1,n_2}} \) is the excess returns of the \( n \) assets in the sense that the first \( n_1 \) elements of \( \tilde{r} - r_{1_{n_1,n_2}} \) are the excess returns of the \( n_1 \) ordinary assets, and the last \( n_2 \) elements of \( \tilde{r} - r_{1_{n_1,n_2}} \) are the returns of the futures contracts, which can be viewed as “excess returns” because futures returns are the payoffs of zero-cost portfolios, just as are the excess returns of ordinary assets.
market portfolio $r_M - r$:

$$\tilde{r} - r_{1_{n_1,n_2}} = \rho W (\tilde{r} - r_{1_{n_1,n_2}}) + \tilde{\alpha} + \beta (r_M - r) + \tilde{\epsilon},$$

$$E[\tilde{\epsilon}] = 0, \quad Cov(r_M, \tilde{\epsilon}) = 0.$$  \hspace{1cm} (2.10)

Then, the S-CAPM implies that, in the above model,

$$\tilde{\alpha} = 0.$$ \hspace{1cm} (2.11)

To see this, rewrite (2.10) as $(I_n - \rho W)(\tilde{r} - r_{1_{n_1,n_2}}) = \tilde{\alpha} + \beta (r_M - r) + \tilde{\epsilon}$. Taking covariance with $r_M$ on both sides and using $Cov(r_M, \tilde{\epsilon}) = 0$ yields $\beta = (I_n - \rho W) \frac{Cov(\tilde{r}, r_M)}{Var(r_M)}$, from which it follows that $\tilde{\alpha} = (I_n - \rho W) E[(\tilde{r} - r_{1_{n_1,n_2}}) - \frac{Cov(\tilde{r}, r_M)}{Var(r_M)} (r_M - r)]$. If the S-CAPM holds, then (2.9) implies $\tilde{\alpha} = 0$.  \footnote{A spatial lag CAPM equation, which is similar to (2.10) with $\tilde{\alpha} = 0$ and considers only ordinary assets but not futures, is defined in Fernandez (2011) without theoretical justification. In contrast, this chapter rigorously proves that the S-CAPM relation (2.9) holds (for both ordinary assets and futures) and that $\tilde{\alpha}$ must be 0 in the spatial model (2.10) under the assumption in Theorem 2.3.1.}

### 2.4 The Spatial Arbitrage Pricing Theory

Unlike the CAPM, the Arbitrage Pricing Theory (APT) introduced by Ross (1976a,b) is based on an asymptotic arbitrage argument rather than on market equilibrium; it allows for multiple risk factors and does not require the identification of the market portfolio. APT provides a linear relationship between the expected asset returns and the factor loadings.  \footnote{There is a vast literature on APT; see e.g., Huberman (1982); Chamberlain (1983); Chamberlain and Rothschild (1983); Ingersoll (1984); Huberman and Wang (2008), among others.}


In this section, we derive the Spatial Arbitrage Pricing Theory (S-APT) and point out its implications. As in Section 2.2, we consider a one-period model with \( n \) risky assets.

Consider the following factor model with spatial interaction:

\[
    r_i = \rho \sum_{j=1}^{n} w_{ij} r_j + \alpha_i + \sum_{k=1}^{K} \beta_{ik} f_k + \epsilon_i, \quad i = 1, \ldots, n, \tag{2.12}
\]

where \( r_i, \rho, w_{ij}, \alpha_i, \epsilon_i \) have the same meaning as in (2.2); \( f_1, \ldots, f_K \) are \( K \) risk factors with \( E[f_k] = 0 \); and \( \beta_{ik} \) is the loading coefficient of the asset \( i \) on the factor \( k \). Let \( \tilde{r} := (r_1, \ldots, r_n)' \), \( W := (w_{ij}) \), \( \alpha := (\alpha_1, \ldots, \alpha_n)' \), \( B := (\beta_{ik}) \), \( \tilde{f} := (f_1, \ldots, f_K)' \), and \( \tilde{\epsilon} := (\epsilon_1, \ldots, \epsilon_n)' \). Then, the above model can be represented in a vector-matrix form as

\[
    \tilde{r} = \rho W \tilde{r} + \alpha + B \tilde{f} + \tilde{\epsilon}, \quad E[\tilde{f}] = 0, \quad E[\tilde{\epsilon}] = 0, \quad E[\tilde{\epsilon}' \tilde{\epsilon}] = V, \quad E[\tilde{f} \tilde{\epsilon}'] = 0. \tag{2.13}
\]

The model (2.13) reduces to the classical APT when \( \rho = 0 \).

### 2.4.1 Asymptotic Arbitrage

We first introduce the notion of asymptotic arbitrage defined in Huberman (1982) and in Ingersoll (1984). Suppose the set of factors \( \tilde{f} = (f_1, \ldots, f_K)' \) are fixed and consider a sequence of economies with increasing numbers of risky assets whose returns depend on these factors and on spatial interaction. As in Section 2.3, in the \( n \)th economy there are \( n_1 \) ordinary assets and \( n_2 \) futures contracts, where \( n = n_1 + n_2 \). Suppose the futures prices of the \( i \)th futures contract are \( F_{i,0}^{(n)} \) and \( F_{i,1}^{(n)} \) at time 0 and time 1, respectively. As in Section
2.3, we define the futures returns as

\[ r^{(n)}_{1+i} = \frac{F^{(n)}_{i+1} - F^{(n)}_{i}}{F^{(n)}_{i}}, \quad i = 1, \ldots, n_2. \]  

(2.14)

Assume the returns \( \tilde{r}^{(n)} = (r_1^{(n)}, \ldots, r_n^{(n)})' \) are generated by

\[
\tilde{r}^{(n)} = \rho^{(n)} W^{(n)} \tilde{r}^{(n)} + \alpha^{(n)} + B^{(n)} \tilde{f} + \tilde{\epsilon}^{(n)}, \quad \text{where}

E[\tilde{f}] = 0, \quad E[\tilde{\epsilon}^{(n)}] = 0, \quad E[\tilde{\epsilon}^{(n)}(\tilde{\epsilon}^{(n)})'] = V^{(n)}, \quad E[\tilde{f} \cdot (\tilde{\epsilon}^{(n)})'] = 0.
\]

(2.15)

The \((n+1)\)th economy includes all the \(n\) risky assets in the \(n\)th economy and one extra risky asset. In the \(n\)th economy, a portfolio is denoted by a vector of dollar-valued positions \( h^{(n)} := (h_1^{(n)}, \ldots, h_{n_1}^{(n)}, h_{n_1+1}^{(n)}, \ldots, h_n^{(n)})' \), where \( h_1^{(n)}, \ldots, h_{n_1}^{(n)} \) denote the dollar-valued wealth invested in the first \(n_1\) assets; \( h_{n_1+i}^{(n)} := O_i F^{(n)}_{i,0} \), where \( O_i \) denotes the number of \(i\)th futures contracts held in the portfolio, and \( i = 1, \ldots, n_2 \). A portfolio \( h^{(n)} \) is a zero-cost portfolio if \( (h^{(n)})'1_{n_1,n_2} = 0 \), where \( 1_{n_1,n_2} \) is defined in (2.8) and the payoff of the zero-cost portfolio is \( (h^{(n)})'(\tilde{r}^{(n)} + 1_{n_1,n_2}) = (h^{(n)})'\tilde{r}^{(n)}.\)

Asymptotic arbitrage is defined to be the existence of a subsequence of zero-cost portfolios \( \{h^{(mk)}_k, k = 1, 2, \ldots\} \) and \( \delta > 0 \) such that

\[
E[(h^{(mk)}_k)'\tilde{r}^{(mk)}_k] \geq \delta, \quad \text{for all} \ k, \quad \text{and} \quad \lim_{k \to \infty} Var((h^{(mk)}_k)'\tilde{r}^{(mk)}_k) = 0. \]

(2.16)

\(^6\)If there is a risk free asset with return \(r\), then a zero-cost portfolio with dollar-valued positions \( h^{(n)} \) in the risky assets must have a dollar-valued position \(-(h^{(n)})'1_{n_1,n_2}\) in the risk free asset. Then, the payoff of the portfolio is given by \( (h^{(n)})'(\tilde{r} - r1_{n_1,n_2}) \).

\(^7\)In the case when there is a risk free asset with return \(r\), the term \( (h^{(mk)}_k)'\tilde{r}^{(mk)}_k \) should be replaced by \( (h^{(mk)}_k)'(\tilde{r}^{(mk)}_k - r1_{n_1,n_2}) \).
2.4.2 The Spatial Arbitrage Pricing Theory: A Special Case in Which Factors Are Tradable

To obtain a good intuition, we first develop the S-APT in the case in which the factors are the payoff of tradable zero-cost portfolios. More precisely, suppose that there is a risk-free return \( r \) and the risk factors \( \tilde{f} \) are given by

\[
\tilde{f} = \tilde{g} - E[\tilde{g}],
\]

(2.17)

where \( \tilde{g} = (g_1, g_2, \ldots, g_K)' \) and each \( g_k \) is the payoff of a certain tradable zero-cost portfolio. The model (2.15) can then be written as

\[
\bar{\tilde{r}}^{(n)} - (\bar{r}_1^{(n)})_{n_1,n_2} = \rho^{(n)} W^{(n)} (\bar{\tilde{r}}^{(n)} - r1_{n_1,n_2}) + \bar{\alpha}^{(n)} + B^{(n)} \tilde{g} + \tilde{\epsilon}^{(n)},
\]

(2.18)

\[
\bar{\alpha}^{(n)} := \alpha^{(n)} - (I_n - \rho^{(n)} W^{(n)}) 1_{n_1,n_2} r - B^{(n)} E[\tilde{g}].
\]

(2.19)

**Theorem 2.4.1** Suppose there is a risk-free return \( r \) and the risk factors \( \tilde{f} \) are given by (2.17) where \( g_1, g_2, \ldots, g_K \) are the payoffs of certain zero-cost portfolios. Suppose

\[
E[\epsilon_i^{(n)} \epsilon_j^{(n)}] = 0, \text{ for } i \neq j; \ Var(\epsilon_i^{(n)}) \leq \sigma^2, \text{ for all } i \text{ and } n,
\]

(2.20)

where \( \sigma^2 \) is a fixed positive number. If there is no asymptotic arbitrage, then

\[
\bar{\alpha}^{(n)} \approx 0,
\]

(2.21)
or, equivalently,

\[ \alpha^{(n)} \approx (I_n - \rho^{(n)}W^{(n)})1_{n_1,n_2}r + B^{(n)}E[\tilde{g}]. \]  

(2.22)

The approximation (2.21) holds in the sense that for any \( \delta > 0 \) there exists a constant \( N_\delta > 0 \) such that \( N(n, \delta) < N_\delta \) for all \( n \), where \( N(n, \delta) \) denotes the number of components of \( \tilde{\alpha}^{(n)} \) whose absolute values are greater than \( \delta \).

**Proof.** See Appendix A.3.1.

The intuition behind the theorem is that if \( \tilde{g} \) are the payoffs of zero-cost portfolios, then, by (2.18), one can construct zero-cost portfolios with payoffs \( \tilde{\alpha}^{(n)} + \tilde{\epsilon}^{(n)} \) that do not carry systematic risk. If the elements of \( \tilde{\epsilon}^{(n)} \) are uncorrelated and have bounded variance, then \( \tilde{\alpha}^{(n)} \) must be approximately zero; otherwise, one could construct a large zero-cost portfolio with a payoff whose mean would be strictly positive while its variance would vanish, constituting an asymptotic arbitrage opportunity.

### 2.4.3 The Spatial Arbitrage Pricing Theory: A General Case

**Theorem 2.4.2** *(S-APT with Both Ordinary Assets and Futures)* Suppose that in the \( n \)th economy there are \( n_1 \) ordinary risky assets and \( n_2 \) futures contracts and the \( n_1 \) ordinary asset returns and the \( n_2 \) futures returns are generated by the model (2.15). If there is no asymptotic arbitrage opportunity, then there is a sequence of factor premiums \( \lambda^{(n)} = \)
$(\lambda_1^{(n)}, \ldots, \lambda_K^{(n)})'$ and a constant $\lambda_0^{(n)}$, which price all assets approximately:

$$\alpha^{(n)} \approx (I_n - \rho^{(n)}W^{(n)})1_{n_1,n_2}\lambda_0^{(n)} + B^{(n)}\lambda^{(n)}. \quad (2.23)$$

The precise meaning of the approximation in (2.23) is that there exists a positive number $A$ such that the weighted sum of the squared pricing errors is uniformly bounded,

$$(U^{(n)})'(V^{(n)})^{-1}U^{(n)} \leq A < \infty \text{ for all } n, \quad (2.24)$$

where

$$U^{(n)} = \alpha^{(n)} - (I_n - \rho^{(n)}W^{(n)})1_{n_1,n_2}\lambda_0^{(n)} - B^{(n)}\lambda^{(n)}.$$ 

In particular, if there exists a risk-free return $r$, then $\lambda_0^{(n)}$ can be identified as $r$.

\textbf{Proof.} See Appendix A.3.2. $\square$

Comparing (2.22) and (2.23), one can see that the factor risk premiums $\lambda^{(n)}$ in the S-APT can be identified as

$$\lambda^{(n)} = E[\tilde{g}] \quad (2.25)$$

if $\tilde{f} = \tilde{g} - E[\tilde{g}]$ and $\tilde{g}$ are the payoffs of zero-cost traded portfolios. The S-APT implies that the degree of spatial interaction affects asset risk premiums. Indeed, let $(\beta_{i,1}^{(n)}, \beta_{i,2}^{(n)}, \ldots, \beta_{i,K}^{(n)})$ be the $i$th row of $(I_n - \rho^{(n)}W^{(n)})^{-1}B^{(n)}$. Then, (2.23) implies that the ordinary assets are
approximately priced by

\[
E[r_i^{(n)}] - \lambda_0^{(n)} \approx \sum_{k=1}^{K} \bar{\beta}_{i,k}^{(n)} \lambda_k^{(n)}, i = 1, \ldots, n_1 
\]  
(2.26)

and that the futures contracts are approximately priced by

\[
\frac{E[F_{i,1}^{(n)}] - F_{i,0}^{(n)}}{F_{i,0}^{(n)}} \approx \sum_{k=1}^{K} \bar{\beta}_{i+1,k}^{(n)} \lambda_k^{(n)}, i = 1, \ldots, n_2. 
\]  
(2.27)

(2.26) and (2.27) show that the expected returns of both ordinary assets and futures contracts are affected by the spatial interaction parameter \( \rho \) because, for all \( j \) and \( k \), \( \bar{\beta}_{j,k}^{(n)} \) depends on the spatial interaction terms \( \rho^{(n)} \) and \( W^{(n)} \).

### 2.4.4 Comparison with the SAR Model

The spatial autoregressive (SAR) model (see, e.g., Lesage and Pace, 2009, Chapter 2.6) is one of the most commonly adopted models in the spatial econometrics literature. The SAR model postulates that the dependent variables (usually prices or log prices of assets) \( y_1, \ldots, y_n \) are generated by

\[
y_i = \rho \sum_{j=1}^{n} w_{ij} y_j + \beta_0 + \sum_{k=1}^{K} \beta_k x_{ik} + \epsilon_i, i = 1, \ldots, n, 
\]  
(2.28)

where \( \rho, w_{ij}, \) and \( \epsilon_i \) have the same meaning as in (2.1); \( x_{ik} \) are explanatory variables; \( \beta_0 \) is the intercept; and \( \beta_1, \ldots, \beta_K \) are coefficients in front of explanatory variables.
Although the first term in (2.28) of the SAR model is the same as the first term of the S-APT model (2.12), there are substantial differences between the two: (i) In terms of model specification, the S-APT imposes a linear constraint on model parameters ((2.22) or (2.23)), while the parameters in the SAR model are free parameters. (ii) In the SAR model (2.28), factors \( x_{ik} \) may be different for different \( i \) and the intercept \( \beta_0 \) is the same for different \( i \); in contrast, in the S-APT model (2.12), factors \( f_k \) are the same for different \( i \) but the intercepts \( \alpha_i \) depend on \( i \). (iii) In terms of parameter estimation, the asymptotic properties of the parameter estimators of the SAR model are obtained when \( n \to \infty \), but those of the S-APT are derived when the number of periods of observation \( T \to \infty \), with \( n \) fixed (see Section 2.5.2).

The difference represented in (iii) is important because, for the SAR model, the MLE may not have the desired \( \sqrt{n} \)-rate of convergence for some specifications of spatial weight matrix \( W \); see Lee (2004). In contrast, the MLE of the S-APT has a \( \sqrt{T} \)-rate of convergence (see Theorem 2.5.2).

2.5 Statistical Inference for the Spatial Arbitrage Pricing Theory

Let \( \tilde{\mathbf{r}}_t = (r_{1t}, r_{2t}, \ldots, r_{nt})' \) be the observation of \( n = n_1 + n_2 \) asset returns that consist of \( n_1 \) ordinary asset returns and \( n_2 \) futures returns in the \( t \)th period. Let \( r_{ft} \) be the risk free return in the \( t \)th period. Let \( \tilde{\mathbf{y}}_t = (y_{1t}, y_{2t}, \ldots, y_{nt})' := \tilde{\mathbf{r}}_t - r_{ft}1_{n_1, n_2} \) denote the excess asset
returns. Let \( \tilde{g}_t = (g_{1t}, g_{2t}, \ldots, g_{Kt})' \) be the observation of the \( k \) factors in the \( t \)th period (note that \( E[\tilde{g}_t] \) may not be zero).

Assume \( \tilde{y}_t \) and \( \tilde{g}_t \) are generated by the following panel data model, a multi-period version of the model (2.18):

\[
\tilde{y}_t = \rho W \tilde{y}_t + \alpha + B \tilde{g}_t + \tilde{\epsilon}_t, \; t = 1, 2, \ldots, T,
\]

\[
(\tilde{y}_t, \tilde{g}_t), \; t = 1, 2, \ldots, T, \; \text{are i.i.d.,}
\]

\[
\tilde{\epsilon}_t \mid \tilde{g}_t \sim N(0, \sigma^2 I_n).
\]

The model (2.29) incorporates three features: (i) a spatial lag in the dependent variables, (ii) individual-specific fixed effects, and (iii) heterogeneity of factor loadings on common factors. However, existing models have as yet incorporated only some but not all of these features.\(^8\)

We need the following mild assumptions regarding \( \tilde{g}_t \) for proving Proposition 2.5.1 and Proposition 2.5.2:

**Assumption 2.5.1** We assume that \( \tilde{g}_t \) satisfies the following mild technical conditions:

\( i \) \( E[\|\tilde{g}_t\|^2] < \infty, \text{ where } \|\tilde{g}_t\|^2 := \sum_{i=1}^{K} g_{it}^2. \)

\(^8\)Lee and Yu (2010a) investigate the asymptotic properties of the QMLEs for spatial panel data models that incorporate the features (i) and (ii) but not (iii); Holly, Pesaran, and Yamagata (2010) and Pesaran and Tosetti (2011) consider panel data models that incorporate spatially correlated cross-section errors and the features (ii) and (iii), but not (i); see Anselin, Le Gallo, and Jayet (2008) and Lee and Yu (2010b) for more comprehensive discussion of spatial panel data models and the asymptotic properties of MLE and QMLE for these models.
(ii) There exists an open set \( \mathbb{A} \subset \mathbb{R}^K \) such that \( P(\tilde{g}_t \in \mathbb{A}) > 0 \) and the distribution of \( \tilde{g}_t \) restricted on \( \mathbb{A} \) has a strictly positive density.

For brevity of notation, we define

\[
b := (\bar{\alpha}_1, \beta_{11}, \beta_{12}, \ldots, \beta_{1K}, \ldots, \bar{\alpha}_n, \beta_{n1}, \beta_{n2}, \ldots, \beta_{nK})'
\]  

(2.30)

and denote the parameter vector of the model as \( \theta := (\rho, b', \sigma^2)' \). Let \( \theta_0 = (\rho_0, b_0', \sigma_0^2)' \) be the true model parameters that lie in the interior of the parameter space \( \Theta \) defined by:

\[
\Theta := [\zeta, \gamma] \times [-\delta_b, \delta_b]^{n \times (K+1)} \times [\delta_s^{-1}, \delta_s],
\]  

(2.31)

where \( \zeta < 0 < \gamma, \delta_b > 0, \delta_s > 0 \) are constants and \( I_n - \rho W \) is invertible for \( \rho \in [\zeta, \gamma] \).

We assume that the spatial weight matrix \( W \) satisfies the following standard conditions:

**Assumption 2.5.2** \( W \) is non-negative; \( W \neq 0 \); and the diagonal elements of \( W \) are all equal to zero.

In the following, we first study the identifiability of the model and then develop the statistical procedures to estimate model parameters and to test the S-APT implication (2.21).

---

9 For any \( W \), because \( \lim_{\rho \to 0} \det(I_n - \rho W) = 1 \), there always exists an interval [\( \zeta, \gamma \)] such that it contains 0 and that \( I_n - \rho W \) is invertible for \( \rho \in [\zeta, \gamma] \). In fact, \( I_n - \rho W \) is invertible if and only if \( \rho^{-1} \) is not an eigenvalue of \( W \) (see footnote 1). Hence, the specification of [\( \zeta, \gamma \)] depends on \( W \): (i) if \( W \) has at least two different real eigenvalues and \( \omega_{\text{min}} < 0 < \omega_{\text{max}} \) are the minimum and maximum real eigenvalues, then [\( \zeta, \gamma \)] can be chosen as an interval that lies inside (\( \omega_{\text{min}}^{-1}, \omega_{\text{max}}^{-1} \)); (ii) if \( W \) does not have real eigenvalues, then [\( \zeta, \gamma \)] can be any interval contains 0. See Lesage and Pace (2009, Chapter 4.3.2, p.88) for more detailed discussion.
2.5.1 Identifiability of the Model

Let

\[ X_t := \begin{pmatrix} 1, g_{1t}, \cdots, g_{Kt} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & 1, g_{1t}, \cdots, g_{Kt} \end{pmatrix} \in \mathbb{R}^{n \times (K+1)}. \quad (2.32) \]

Then, \( X_t \beta = \bar{\alpha} + B \bar{g}_t \) and the log likelihood function of the model is given by

\[
\ell(\theta) = \ell(\rho, \bar{\beta}, \sigma^2) := \sum_{t=1}^{T} l(\tilde{y}_t | \tilde{g}_t, \theta), \quad \text{where}
\]

\[
l(\tilde{y}_t | \tilde{g}_t, \theta) = -\frac{n}{2} \log(2\pi\sigma^2) + \frac{1}{2} \log(\det((I_n - \rho W')(I_n - \rho W)))
- \frac{1}{2\sigma^2}(\tilde{y}_t - \rho W \tilde{\tilde{y}}_t - X_t \bar{\beta})(\tilde{y}_t - \rho W \tilde{\tilde{y}}_t - X_t \bar{\beta})'. \quad (2.34)
\]

We recall the following definition of identifiability; see, e.g., Neway and McFadden (1994, Lemma 2.2).

**Definition 2.5.1** \( \theta_0 \) is identifiable if for any \( \theta \neq \theta_0 \) and \( \theta \in \Theta \) it holds that \( P(l(\tilde{y}_t | \tilde{g}_t, \theta) \neq l(\tilde{y}_t | \tilde{g}_t, \theta_0)) > 0. \)

It turns out that the identifiability of \( \theta_0 \) depends largely on the property of the spatial weight matrix \( W \). In particular, we need the following definition for \( W \):

**Definition 2.5.2** The spatial weight matrix \( W \) is regular if there exist no \( c_1 > 0 \) and \( c_2 \geq 0 \)
such that

\[
\sum_{k=1}^{n} W_{ki}^2 = c_1, \quad \forall i = 1, \ldots, n, \quad (2.35)
\]

\[
\sum_{k=1}^{n} W_{ki} W_{kj} = c_2(W_{ij} + W_{ji}), \quad \forall 1 \leq i < j \leq n. \quad (2.36)
\]

If \( W \) is not regular, then the pair of constants \((c_1, c_2)\) that satisfy (2.35) and (2.36) are unique. Indeed, by Assumption 2.5.2, there exists \( i < j \) such that \( W_{ij} + W_{ji} > 0 \); hence, it follows from (2.36) that \( c_2 \) is uniquely determined by \( W \). Apparently, \( c_1 \) is also unique.

We have the following proposition regarding the identifiability of \( \theta_0 \).

**Proposition 2.5.1** Any \( \theta_0 \) is identifiable if \( W \) is regular. More generally, a particular \( \theta_0 \) is identifiable if and only if \( W \) satisfies one of the following conditions:

(i) \( W \) is regular.

(ii) \( W \) is not regular and corresponds to the unique pair \((c_1, c_2)\) in (2.35) and (2.36), and one of the following conditions holds:

\[
\rho_0 = -\frac{c_2}{c_1}; \quad (2.37)
\]

\[
\rho_0 \neq -\frac{c_2}{c_1} \text{ and } \frac{1 - c_2 \rho_0}{c_2 + c_1 \rho_0} = \rho_0; \quad (2.38)
\]

\[
\rho_0 \neq -\frac{c_2}{c_1} \text{ and } \frac{1 - c_2 \rho_0}{c_2 + c_1 \rho_0} \neq \rho_0 \text{ and } \theta_* := (\rho_*, \bar{\alpha}_*, B_*, \sigma_*^2) \notin \Theta, \quad (2.39)
\]
where \( \rho_* := \frac{1-c_2 \rho_0}{c_2 + c_1 \rho_0} \), \( \sigma_*^2 := \sigma_0^2 \frac{c_2^2 + c_1}{(c_2 + c_1 \rho_0)^2} \), \( \bar{\alpha}_* := \frac{\sigma_*^2}{\sigma_0^2} (I_n - \rho_* W')^{-1} (I_n - \rho W') \bar{\alpha}_0 \), and \( B_* := \frac{\sigma_*^2}{\sigma_0^2} (I_n - \rho_* W')^{-1} (I_n - \rho W') B_0 \).

**Proof.** See Appendix A.4. \( \square \)

Proposition 2.5.1 is equivalent to the following statement: A particular \( \theta_0 \) is not identifiable if and only if \( W \) is not regular and it corresponds to the unique pair \((c_1, c_2)\) with \( c_1 > 0 \) and \( c_2 \geq 0 \) in (2.35) and (2.36), and \( \rho_0 \neq -\frac{c_2}{c_1} \), and \( \rho_0 \neq \frac{1-c_2 \rho_0}{c_2 + c_1 \rho_0} \), and \( \theta_* \in \Theta \). In all the empirical examples of the chapter, \( W \) is regular.\(^{10} \) In the rest of the section, we assume that \( W \) is regular and hence \( \theta_0 \) is identifiable.

### 2.5.2 Model Parameter Estimation

**Theorem 2.5.1** *(Representation of MLE)* The MLE \( \hat{\rho} \) is given by

\[
\hat{\rho} = \arg \max_{\rho \in [\zeta, \gamma]} \ell_c(\rho),
\]

\(^{10} \)By Definition 2.5.2, if \( W \) is not regular, then the elements of \( W \) satisfy \( n(n + 1)/2 \) constraints given by (2.35) and (2.36); hence, unless \( W \) is carefully constructed to satisfy these constraints, \( W \) is regular and the (unknown) true parameter is identifiable. For example, when \( W \) is not regular and \( n = 3 \), \( W \) has six off-diagonal elements that satisfy six constraints; hence, only very special \( W \) are not regular.
where

\[
\ell_c(\rho) := \ell(\rho, b(\rho), s(\rho))
\]

\[
= -\frac{nT}{2} \log(2\pi s(\rho)) + \frac{T}{2} \log(\det((I_n - \rho W')(I_n - \rho W))) - \frac{nT}{2},
\]

(2.40)

\[
s(\rho) := \frac{1}{nT} \sum_{t=1}^{T} ((I_n - \rho W)\tilde{y}_t - X_t b(\rho))'((I_n - \rho W)\tilde{y}_t - X_t b(\rho))
\]

\[
b(\rho) := \left( \sum_{t=1}^{T} X_t'X_t \right)^{-1} \sum_{t=1}^{T} X_t'(I_n - \rho W)\tilde{y}_t.
\]

(2.41)

(2.42)

And the MLE \( \hat{b} \) and \( \hat{\sigma}^2 \) are given by \( \hat{b} = b(\hat{\rho}) \), \( \hat{\sigma}^2 = s(\hat{\rho}) \).

**Proof.** For any given \( \rho \), the original model can be rewritten as

\[
\tilde{y}_t - \rho W\tilde{y}_t = X_t b + \tilde{\epsilon}_t, \quad t = 1, 2, \ldots, T,
\]

from which the classical theory of linear regression shows that \( \hat{b} = b(\rho) \) and \( \sigma^2 = s(\rho) \) maximize the log likelihood function (2.33). Because \( \ell_c(\rho) = \ell(\rho, b(\rho), s(\rho)) \) and \( \hat{\rho} \) maximizes \( \ell_c(\rho) \), it follows that \( \hat{\rho}, \hat{b} = b(\hat{\rho}), \) and \( \hat{\sigma}^2 = s(\hat{\rho}) \) maximize \( \ell(\rho, \hat{b}, \sigma^2) \), i.e., they are the MLE.

We need the following proposition to show the asymptotic properties of the MLE.
Proposition 2.5.2 Define

\[
Q_0(\theta) := E[l(\tilde{y}_t \mid \tilde{g}_t, \theta)], \quad \hat{Q}_T(\theta) := \frac{1}{T} \sum_{t=1}^{T} l(\tilde{y}_t \mid \tilde{g}_t, \theta).
\] (2.43)

Then, \(\hat{Q}_T(\theta)\) is twice continuously differentiable on the interior of \(\Theta\). Define

\[
s(\tilde{y}_t, \tilde{g}_t; \theta) := \frac{\partial l(\tilde{y}_t \mid \tilde{g}_t, \theta)}{\partial \theta} \quad \text{and} \quad H(\tilde{y}_t, \tilde{g}_t; \theta) := \frac{\partial^2 l(\tilde{y}_t \mid \tilde{g}_t, \theta)}{\partial \theta \partial \theta'}.
\] (2.44)

Then, the following statements hold:

(i) \(Q_0(\theta)\) is uniquely maximized at \(\theta_0\).

(ii) \(\sup_{\theta \in \Theta} |\hat{Q}_T(\theta) - Q_0(\theta)| \xrightarrow{p} 0\), as \(T \to \infty\).

(iii) \(Q_0(\theta)\) is continuous on \(\Theta\).

(iv) \(E[s(\tilde{y}_t, \tilde{g}_t; \theta_0)] = 0\).

(v) \(H(\tilde{y}_t, \tilde{g}_t; \theta)\) is equal to

\[
\begin{bmatrix}
\tilde{v} & -\frac{1}{\sigma^2} \tilde{y}_t'W'X_t & -\frac{1}{\sigma^2} \tilde{y}_t'W'\tilde{\xi}_t \\
-\frac{1}{\sigma^2} X_t'W\tilde{y}_t & -\frac{1}{\sigma^2} X_t'X_t & -\frac{1}{\sigma^2} X_t'\tilde{\xi}_t \\
-\frac{1}{\sigma^2} \tilde{\xi}_t'W\tilde{y}_t & -\frac{1}{\sigma^2} \tilde{\xi}_t'X_t & -\frac{n}{2\sigma^2} - \frac{1}{\sigma^4} \tilde{\xi}_t'\tilde{\xi}_t
\end{bmatrix},
\] (2.45)

where \(\tilde{\xi}_t := (I_n - \rho W)\tilde{y}_t - X_t\tilde{v}\), \(\tilde{v} = -\text{tr}(W(I_n - \rho W)^{-1}W(I_n - \rho W)^{-1}) - \frac{1}{\sigma^2} \tilde{y}_t'W'W\tilde{y}_t\) and \(\text{tr}(\cdot)\) denotes the matrix trace.
\( (vi) \quad -E[H(\tilde{y}_t, \tilde{g}_t; \theta_0)] = E[s(\tilde{y}_t, \tilde{g}_t; \theta_0)] s(\tilde{y}_t, \tilde{g}_t; \theta_0)'. \)

\( (vii) \quad E[H(\tilde{y}_t, \tilde{g}_t; \theta_0)] \) is invertible.

\( (viii) \quad \text{There is a neighborhood } \mathcal{N} \text{ of } \theta_0 \text{ such that } E[\sup_{\theta \in \mathcal{N}} \|H(\tilde{y}_t, \tilde{g}_t; \theta)\|] < \infty. \)

\textbf{Proof.} See Appendix A.5.

\textbf{Theorem 2.5.2} \textit{(Asymptotic properties of the MLE)} The MLE \( \hat{\theta} := (\hat{\rho}, \hat{\sigma}, \hat{\sigma}^2) \) has consistency and asymptotic normality:

\[ \begin{align*}
(i) \quad \hat{\theta} & \xrightarrow{p} \theta_0, \quad \text{as } T \to \infty, \quad (2.46) \\
(ii) \quad \sqrt{T}(\hat{\theta} - \theta_0) & \xrightarrow{d} N(0, -E[H(\tilde{y}_t, \tilde{g}_t; \theta_0)]^{-1}), \quad \text{as } T \to \infty, \quad (2.47) \\
(iii) \quad \frac{1}{T} \sum_{t=1}^{T} H(\tilde{y}_t, \tilde{g}_t; \hat{\theta}) & \xrightarrow{p} E[H(\tilde{y}_t, \tilde{g}_t; \theta_0)], \quad \text{as } T \to \infty, \quad (2.48)
\end{align*} \]

where \( H(\tilde{y}_t, \tilde{g}_t; \theta) \) is equal to (2.45).

\textbf{Proof.} By (i), (ii), and (iii) of Proposition 2.5.2 and the compactness of \( \Theta \), it follows from Theorem 2.1 in Neway and McFadden (1994) that consistency (2.46) holds.

We will show the asymptotic normality (2.47) by applying Proposition 7.9 in Hayashi (2000, p. 475). The conditions are verified in the following steps. The consistency of \( \hat{\theta} \) has been proved in above. The condition (1) of Proposition 7.9 holds by the assumption that \( \theta_0 \) lies in the interior of \( \Theta \). The conditions (2), (3), (4), and (5) of Proposition 7.9 follow from Proposition 2.5.2 in this chapter. Hence, all the conditions of Proposition 7.9 hold and its conclusion implies (2.47).
At last, we will show that (2.48) holds. Define 
\[ \hat{H}(\theta) := \frac{1}{T} \sum_{t=1}^{T} H(\tilde{y}_t, \tilde{g}_t; \theta) . \]
Let \( N \) be a neighborhood such that (viii) in Proposition 2.5.2 holds and let \( \Theta_0 \subset N \) be a compact set that contains \( \theta_0 \). Then, it follows from (viii) in Proposition 2.5.2 and Lemma 2.4 in Neway and McFadden (1994, p. 2129) that 
\[ H(\theta) := E[H(\tilde{y}_t, \tilde{g}_t; \theta)] \]
is continuous and
\[ \sup_{\theta \in \Theta_0} \left\| \frac{1}{T} \sum_{t=1}^{T} \frac{\partial^2 l(\tilde{y}_t | \tilde{g}_t, \theta)}{\partial \theta \partial \theta'} - H(\theta) \right\| \overset{p}{\to} 0, \text{ as } T \to \infty. \] (2.49)

Since \( \hat{\theta} \overset{p}{\to} \theta_0 \) and \( H(\theta) \) is continuous, it follows that \( H(\hat{\theta}) \overset{p}{\to} H(\theta_0) \). For any \( \varepsilon > 0 \),
\[
P\left( \| \hat{H}(\hat{\theta}) - H(\theta_0) \| > \varepsilon \right) 
\leq P\left( \| \hat{H}(\hat{\theta}) - H(\hat{\theta}) \| + \| H(\hat{\theta}) - H(\theta_0) \| > \varepsilon \right)
\leq P\left( \| \hat{H}(\hat{\theta}) - H(\hat{\theta}) \| > \frac{\varepsilon}{2}, \hat{\theta} \in \Theta_0 \right) + P\left( \| H(\hat{\theta}) - H(\theta_0) \| > \frac{\varepsilon}{2}, \hat{\theta} \notin \Theta_0 \right) + P\left( \| H(\hat{\theta}) - H(\theta_0) \| > \frac{\varepsilon}{2} \right)
\to 0, \text{ as } T \to \infty,
\]
where the limit follows from (2.49), \( \hat{\theta} \overset{p}{\to} \theta_0 \), and \( H(\hat{\theta}) \overset{p}{\to} H(\theta_0) \).

The asymptotic properties of the MLE of S-APT model (2.29) are obtained by letting \( T \to \infty \) and keeping \( n \) fixed; in contrast, those of the SAR model are obtained by letting \( n \to \infty \). As a result, the MLE of the S-APT model has a \( \sqrt{T} \)-rate of convergence as long as \( W \) satisfies the identifiability condition specified in Proposition (2.5.1), but those of the
SAR may not have the desired $\sqrt{n}$-rate of convergence when $W$ is not sparse enough; see Lee (2004).\textsuperscript{11}

We investigate the finite-sample performance of the estimators using 2000 data sets simulated from the model (2.29). In all of the simulation studies below, we use the locations of twenty major cities in the United States as asset locations and $W$ is defined by the method of Delaunay triangularization, which is commonly adopted in spatial econometrics literature (see, e.g., Pace, 2003).\textsuperscript{12} It is easy to check that the specified matrix $W$ is regular;\textsuperscript{13} hence, by Proposition 2.5.1, the model parameters are identifiable.

We specify $\bar{\alpha}_0 = 0$ and $\sigma_0^2 = 0.5$. An i.i.d. draw of 20 samples from $N(0, 1)$ is fixed as the elements of $B_0$. $\{\tilde{g}_t : t = 1, \ldots, 131\}$ is generated as one realization of 131 i.i.d. random variables with distribution $N(0.5, 0.5)$. For the fixed $B_0$ and $\{\tilde{g}_t : t = 1, \ldots, 131\}$, 2000 i.i.d. samples of $\{\tilde{\epsilon}_t : t = 1, \ldots, 131\}$ are then simulated and $\{\tilde{y}_t : t = 1, \ldots, 131\}$ are then computed from (2.29). Then, the MLE $\hat{\rho}$ is obtained from each of the simulated data sets.

Table 2.1 shows the mean and standard deviation of the MLE $\hat{\rho}$ for the 2000 simulated data sets for different values of $\rho_0 = 0.2, 0.4, 0.6,$ and $0.8$, respectively. Figure 2.2 shows the histogram of the 2000 estimates $\hat{\rho}$ for the different values of $\rho_0$, which seems to indicate that $\hat{\rho}$ has an asymptotic normal distribution with mean being $\rho_0$.

\textsuperscript{11}More precisely, Lee (2004) shows that when each asset can be influenced by many neighbors, various components of the estimators may have different rates of convergence.

\textsuperscript{12}The twenty cities correspond to the twenty MSAs that have S&P/Case-Shiller home price indices; their locations are specified by their geographic coordinates. See Lesage and Pace (2009, Chap. 4.11) for the details of the method of Delaunay triangularization. We use the program fdelw2 in the Spatial Statistics Toolbox (Pace, 2003) to compute $W$ by this method.

\textsuperscript{13}This is because the sum of square of different columns of $W$ are not equal.
Table 2.1: The mean and standard deviation of $\hat{\rho}$. The asymptotic standard deviations are estimated from the sample average of Hessian matrix (see (2.48)).

<table>
<thead>
<tr>
<th>$\rho_0$</th>
<th>0.2</th>
<th>0.4</th>
<th>0.6</th>
<th>0.8</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean of $\hat{\rho}$</td>
<td>0.199</td>
<td>0.399</td>
<td>0.599</td>
<td>0.799</td>
</tr>
<tr>
<td>(theoretical) asymptotic standard deviation of $\hat{\rho}$</td>
<td>0.028</td>
<td>0.025</td>
<td>0.019</td>
<td>0.011</td>
</tr>
<tr>
<td>empirical standard deviation of $\hat{\rho}$</td>
<td>0.031</td>
<td>0.029</td>
<td>0.020</td>
<td>0.013</td>
</tr>
</tbody>
</table>

Figure 2.2: Histogram of the MLE $\hat{\rho}$ for 2000 data sets simulated from the model (2.29) for different values of $\rho_0$ with $n = 20$, $K = 1$, and $T = 131$. 

32
2.5.3 Hypothesis Test of the S-APT

For simplicity, we assume that the factors $\tilde{g}$ are the payoffs of zero-cost tradable portfolios. In this case, Theorem 2.4.1 shows that the S-APT imposes an approximate constraint $\bar{\alpha}^{(n)} \approx 0$ (see (2.21)). As in the classical factor pricing literature, we test the S-APT by testing the exact constraint

$$
H_0 : \bar{\alpha} = 0; \ H_1 : \bar{\alpha} \neq 0,
$$

(2.50)

where $\bar{\alpha}$ is defined in the model (2.29).

**Theorem 2.5.3** Under the null hypothesis, the likelihood ratio test statistic

$$
LR = 2 \left[ \sum_{t=1}^{T} l(\tilde{y}_t \mid \tilde{g}_t, \hat{\theta}) - \sum_{t=1}^{T} l(\tilde{y}_t \mid \tilde{g}_t, \theta^*) \right]
$$

(2.51)

has an asymptotic $\chi^2(n)$ distribution. Here, $\sum_{t=1}^{T} l(\tilde{y}_t \mid \tilde{g}_t, \hat{\theta})$ denotes the log likelihood function evaluated at $\hat{\theta}$, which is the MLE of parameters estimated with no constraints; while $\sum_{t=1}^{T} l(\tilde{y}_t \mid \tilde{g}_t, \theta^*)$ its counterpart evaluated at the MLE $\theta^*$ estimated under the constraint that the null holds.

**Proof.** We prove the theorem by verifying the conditions of Proposition 7.11 in Hayashi (2000, p. 494), which is concerned with testing the hypothesis that $a(\theta) = b$ for any function $a(\theta)$ of the model parameter $\theta$. In our problem, the function $a(\theta) = \bar{\alpha}$ (as in Equation (2.29)) and the constant $b = 0$. It can be seen that the Jacobian $\frac{\partial a(\theta)}{\partial \theta}$ is of
full row rank. It then suffices to verify the conditions of Proposition 7.9 in Hayashi (2000, p. 475), but it is done in the proof of Theorem 2.5.2 in this chapter.

We carry out the test for 10000 simulated data sets at the confidence level of 95%.\textsuperscript{14} The size of the test is 5.91%, which is slightly higher than the theoretical value of 5%. This may result from small sample bias, as discussed in Campbell, Lo, and MacKinlay (1996, Chap. 5.4).

### 2.5.4 Goodness of Fit of the Model (2.29)

The adjusted $R^2$ for the $i$th asset in the model (2.29) is defined as

$$R^2_i = 1 - \frac{T - 1}{T - K - 1} \frac{\text{Var}(\epsilon_i)}{\text{Var}(y_i)}, \quad i = 1, 2, \ldots, n,$$

where $\epsilon_i$ are the residuals and $y_i$ are the realized returns.

To show the effectiveness of the adjusted $R^2$, two data sets are simulated according to the same model specification as that in Table 2.1, except that $\rho_0$ is fixed at 0.5, and two values of $\sigma_0^2$ (0.01 and 0.5) are used, respectively, for the two data sets, which correspond to the two cases of high and low adjusted $R^2$. In the simulation, first, the factor realization $\{\tilde{g}_t, t = 1, \ldots, T\}$ is simulated and fixed. Then, for each chosen value of $\sigma_0^2$, the residuals $\{\tilde{\epsilon}_t, t = 1, \ldots, T\}$ are simulated and the realized returns $\{\tilde{y}_t, t = 1, \ldots, T\}$ are generated according to the model (2.29). For each simulated data set, we calculate the MLE estimate

\textsuperscript{14}The test data include the 8000 data sets used for Table 2.1 and additional 2000 data sets simulated in the same way with $\rho_0 = 0.5$. 
\( \theta^* \) under the constraint \( \bar{\alpha} = 0 \) and obtain the fitted residual series \( \{ \hat{\epsilon}_t = \bar{y}_t - \rho^* W \bar{y}_t - B^* \tilde{g}_t : t = 1, \ldots, T \} \), where \( \rho^* \) and \( B^* \) are the MLE. The sample adjusted \( R^2 \) of \( y_t \) is computed and compared to the theoretical adjusted \( R^2 \) of \( y_t \).\(^{15}\) Table 2.2 shows that the sample adjusted \( R^2 \) and the theoretical adjusted \( R^2 \) align well.

\[
\begin{array}{cccccccccc}
\sigma_0^2 = 0.01 & & & & & & & & & \\
\hline
& r_1 & r_2 & r_3 & r_4 & r_5 & r_6 & r_7 & r_8 & r_9 & r_{10} \\
\hline
\text{theoretical adjusted } R^2 & 0.7763 & 0.9818 & 0.0910 & 0.4064 & 0.9654 & 0.9730 & 0.9690 & 0.3210 & 0.8415 & 0.3110 \\
\text{sample adjusted } R^2 & 0.7748 & 0.9817 & 0.0848 & 0.4023 & 0.9652 & 0.9728 & 0.9688 & 0.3162 & 0.8404 & 0.3062 \\
\hline
\sigma_0^2 = 0.5 & & & & & & & & & \\
\hline
& r_{11} & r_{12} & r_{13} & r_{14} & r_{15} & r_{16} & r_{17} & r_{18} & r_{19} & r_{20} \\
\hline
\text{theoretical adjusted } R^2 & 0.1833 & 0.9203 & 0.8952 & 0.9906 & 0.6367 & 0.6634 & 0.9672 & 0.0451 & 0.2236 & 0.9394 \\
\text{sample adjusted } R^2 & 0.1777 & 0.9197 & 0.8945 & 0.9905 & 0.6342 & 0.6611 & 0.9670 & 0.0385 & 0.2183 & 0.9390 \\
\hline
\end{array}
\]

Table 2.2: Simulation study of the sample adjusted \( R^2 \). We use the same model specification as that for Table 2.1, except that \( \rho_0 \) is fixed at 0.5 and two values of \( \sigma_0^2 \) (0.01 and 0.5) are used, respectively, for the two data sets. For each data set, the MLE of parameters is estimated for the model (2.29) under the constraint \( \bar{\alpha} = 0 \) and then the sample adjusted \( R^2 \) for each element of \( \tilde{r} \) is calculated and compared to its theoretical counterpart. It appears that the sample adjusted \( R^2 \) and its theoretical counterpart align well.

\(^{15}\)The theoretical adjusted \( R^2 \) of \( y_i \) is calculated using (2.52), where \( \text{Var}(\epsilon_i) = \sigma_0^2 \) and \( \text{Var}(y_i) \) is equal to the \( i \)th diagonal element of the covariance matrix \( (I_n - \rho_0 W)^{-1} B_0 \cdot \text{Cov}(\hat{g}) \cdot B_0^\prime (I_n - \rho_0 W')^{-1} + \sigma_0^2 (I_n - \rho_0 W)^{-1} (I_n - \rho_0 W')^{-1} \). The sample adjusted \( R^2 \) of \( y_i \) is calculated using (2.52) with \( \text{Var}(\epsilon_i) \) and \( \text{Var}(y_i) \) replaced by their sample counterparts.
2.6 Empirical Study of the CSI Indices Futures Using the S-APT

2.6.1 The Data

The CSI Indices are constructed based on the method proposed by Case and Shiller (1987) and are the leading measure of single family home prices in the United States. The CSI index family includes twenty indices for twenty metropolitan areas and three composite indices (National, 10-City, and 20-City). The indices are updated monthly, except for the national index, which is updated quarterly. The CSI Indices themselves are not directly traded; however, CSI Indices futures are traded on the Chicago Mercantile Exchange. There are, in total, eleven CSI Indices futures contracts; one is on the composite 10-City CSI Index and the other ten are on the CSI Indices of ten metropolitan areas: Boston, Chicago, Denver, Las Vegas, Los Angeles, Miami, New York, San Diego, San Francisco, and Washington, D.C.

On any given day, the futures contract with the nearest maturity among all the traded futures contracts is called the first nearest-to-maturity contract. In the empirical study, we use the prices of first nearest-to-maturity futures contract to define one-month ahead nominal return of futures because this contract usually has better liquidity than the others. The time period for the monthly futures return data ranges from June 2006 to July 2011.
2.6.2 Empirical Results

We divide the 10 CSI Indices futures into the west-coast group (Denver, Las Vegas, Los Angeles, San Diego, and San Francisco) and the east-coast group (Boston, Chicago, Miami, New York, and Washington, D. C.). For each group, we carry out the empirical study in three respects: (i) we fit the model (2.29) to the data; (ii) we perform the S-APT hypothesis test (2.50); (iii) we check the goodness-of-fit of the model with the S-APT constraint by inspecting the adjusted $R^2$.

As in the simulation studies, the spatial weight matrix $W$ is specified using Delaunay triangularization for each of the two groups. The two $W$ are regular because the sums of the square of elements of different columns are different; hence, the models are identifiable. For each group, we fit a single-factor model (2.29) to the futures return data. The single factor for a group is defined as the payoff of the ex-post mean-variance efficient portfolio composed of the CSI Indices futures in that group. For instance, let $\tilde{r} = (r_1, \ldots, r_5)'$ be the returns of the five CSI Indices futures in the west-coast group and $\phi = (\phi_1, \ldots, \phi_5)'$ be the dollar-valued positions of a portfolio consisting of the five futures contracts. We solve for the optimal solution $\phi^*$ to the following mean-variance problem with some target mean payoff $e$:

$$\min_{\phi} \frac{1}{2} \phi' \hat{\Sigma} \phi \quad s.t. \quad \phi' \hat{\mu} = e,$$

$$\hat{\Sigma} \quad \hat{\mu}$$

where $\hat{\Sigma}$ and $\hat{\mu}$ are the sample covariance and sample mean (in sample, i.e. from June 2006
to July 2011) of $\tilde{r}$, respectively. We then define $g := (\phi^+)'\tilde{r}$ to be the single factor for the west-coast group.

The MLE of model parameters are obtained by maximizing $\ell_c(\rho)$ defined in (2.40) numerically by comparing function values evaluated at $\rho = k\Delta$ with $\Delta = 0.0001$ and $k$ being integers.

With the single factors corresponding to $e = 2.5\%$, the S-APT constraint is not rejected with p-values of 0.65 and 0.99 for the west-coast and east-coast groups, respectively. Furthermore, the estimation results show that the spatial interaction parameter $\rho$ is significantly positive. Indeed, $\rho$ is estimated to be 0.42 and 0.48 with 95% confidence intervals $[0.32, 0.52]$ and $[0.37, 0.56]$ for the west-coast and east-coast groups, respectively.

Figure 2.6.2 shows the sample adjusted $R^2$ of fitting the one-factor models with the S-APT constraint $\bar{\alpha} = 0$ to the two groups of futures returns. All of the sample adjusted $R^2$ are positive except that of New York, which is $-0.68$. The negative adjusted $R^2$ may be due to the fact that the CSI Index for New York does not reflect the overall real estate market in that city, as it takes into account only single-family home prices but not co-op or condominium prices; however, sales of co-ops and condominiums account for 98% of Manhattan’s non-rental properties.\footnote{We try to alleviate the problem by including a condominium index return factor but this does not improve the fitting results much. As there are no futures contracts on the S&P/Case-Shiller Condominium Index of New York, we construct a mimicking portfolio of the excess return of the Condominium Index using the linear projection of the Condominium Index excess return on the payoff space spanned by the ten CSI Indices futures returns. Then, the payoff of the mimicking portfolio is defined as an additional factor. However, the sample adjusted $R^2$ of the linear projection is merely 17%, indicating that the mimicking portfolio payoff may not be a good approximation to the Condominium Index excess return.} Therefore, we exclude the CSI Indices futures of New York from the east-coast group and test the S-APT on the remaining four CSI Indices.
futures in the east-coast group. The p-value of the test is 0.86, and hence the S-APT is not rejected. The sample adjusted $R^2$ of fitting the remaining four CSI Indices futures under the S-APT constraint is shown in the bottom of Figure 2.3; all four futures have positive adjusted $R^2$; $\rho$ is estimated to be 0.48 with the 95% confidence interval [0.37, 0.56], which is statistically significant.

![Figure 2.3: Adjusted $R^2$ of model fitting. The figure at the top shows the adjusted $R^2$ of fitting the one-factor model (2.29) with the S-APT constraint $\bar{\alpha} = 0$ to the west-coast and east-coast groups of CSI Indices futures returns, respectively; the payoff of the mean-variance efficient portfolio of the futures in the west-coast (east-coast) group is computed by solving (2.53) with $e = 2.5\%$ and is used as the single factor for the west-coast (east-coast) group. $W$ is specified using Delaunay triangulation. The figure at the bottom shows the adjusted $R^2$ of the same model fitting as in the top figure, except that New York is excluded from the east-coast group.](image-url)
2.6.3 Robustness

The empirical results reported above seem to be robust with respect to different specifications of spatial matrix $W$. In part (a) of Table 2.3 we compare the estimation and test results using three definitions of $W$: (i) $W_{ij} := (s_i d_{ij})^{-1}$ where $d_{ij}$ is the geographic distance\(^{17}\) between the $i$th and the $j$th MSA and $s_i := \sum_{j \neq i} d_{ij}^{-1}$; (ii) $W_{ij} := (s_i d_{ij})^{-1}$ and $d_{ij}$ is the driving distance; (iii) $W$ specified by Delaunay triangularization. It is clear that for each group and for each specification of $W$, the p-values for testing $\bar{\alpha} = 0$ are high enough so that S-APT is not rejected; in addition, $\rho$ is significantly positive in all cases.

The test results of the S-APT constraint also seem to be robust with respect to the choice of $e$ in (2.53), as is illustrated in part (b) of Table 2.3. The p-values are high and hence the S-APT constraint is not rejected in any of the cases.

2.7 Conclusion

Real estate assets are distinct from other financial assets in that spatial interaction plays a critical role in determining prices and returns. Although there have been some empirical studies of housing prices using spatial econometric models, there is as yet little work that studies the explicit implications of spatial interaction for expected returns of real estate securities in market equilibrium or under the condition of absence of arbitrage.

In this chapter, we attempt to fill this gap by studying how spatial interaction affects the

\(^{17}\)The geographic distance is calculated from the longitude and latitude coordinates using the Vincenty’s formulae (Vincenty, 1975), which assumes that the figure of the earth is an oblate spheroid instead of a sphere.
relationship between risk and return of real estate securities. We propose new asset pricing models that incorporate spatial interaction, i.e., the spatial capital asset pricing model (S-CAPM) and the spatial arbitrage pricing theory (S-APT), which extend the classical asset pricing theory of CAPM and APT. The S-CAPM and S-APT explicitly characterize the effect of spatial interaction on the expected returns of both ordinary assets and futures contracts. We then develop the econometric tools for implementing the proposed model: (i) We derive the necessary and sufficient condition for the identifiability of parameters. (ii) We derive the MLE and establish the consistency and asymptotic normality of the MLE. In contrast to the MLE for the SAR models, which may not have the desired \( \sqrt{n} \) rate of convergence when \( W \) is not sparse enough, the MLE of the proposed model has a \( \sqrt{T} \) rate of convergence with \( T \) being the number of time periods of observation. (iii) We develop
the likelihood ratio test statistic for testing the S-APT. Finally, an empirical study of the futures contracts on S&P/Case-Shiller Home Price Indices shows that the S-APT is not rejected and that the spatial interaction parameter in the model for the CSI Indices futures returns of ten U.S. cities is statistically significant.
Chapter 3

Jumps in Equity Returns Before and During the Financial Crisis

3.1 Introduction

It is well known that jump risk affects equity returns significantly; see, e.g., Duffie, Pan and Singleton (2000) and Singleton (2006), and the references therein. We attempt to answer two questions about jumps in equity returns: (i) How did jumps in equity returns change during the financial crisis 2007–2011; in particular, were there significant changes in jump rates or in jump sizes, or both? (ii) In any finite-time horizon, were there finite
number of large jumps (e.g. those in affine jump-diffusion models) or infinite number of small jumps (e.g. those in Lévy type models) in equity returns before and during the crisis?

For the first question, the increase of jump rates when market is in distress, especially for the 1987 crash and the tech-bubble burst around 2001–2002, are documented in Pan (2002), Eraker (2004) and Johannes, Polson and Kumar (1999). However, whether there are significant changes in jump-sizes during financial crises has not been addressed in the existing literature. Also the previous empirical studies in general do not distinguish positive and negative jump rates. Based on the latest data on S&P 500 and Nasdaq 100 daily returns up to December 2011, we find both positive and negative jump rates increased significantly during the financial crisis, while, somewhat surprisingly, there is little evidence that average jump sizes have changed before and after the crisis; see Tables 3.8 and 3.11 for S&P 500 and Nasdaq 100, respectively.

The results in the existing literature regarding the second question are mixed. For example, Eraker, Johannes and Polson (2003) find an affine jump-diffusion model with stochastic volatility and correlated jumps in returns and volatility fits S&P 500 data from 1980 to 1999 well, which suggests that there were finite number of moderate jumps (since the tail of normal distribution is not as heavy as the exponential tail). However, Li, Wells and Yu (2008) stress the importance of infinitely many small jumps by fitting a stochastic volatility model with jumps in returns following a variance-gamma process, which is a special case of Lévy process. Li, Wells and Yu (2008) show that the Lévy-type model outperforms the

There could be two plausible explanations for the disparities above. (1) Neither of the two papers mentioned above covers the data related to the current financial crisis, analysis of which could provide new insights in model comparison. In this chapter, we use the data of S&P 500 from 1980 to 2011, and Nasdaq 100 from 2001 to 2011, both covering the 2007-2008 financial crisis. (2) The conclusions in the two papers aforementioned are based on particular modeling assumptions about the jump size distribution. In particular, both papers aforementioned assume that the affine jump-diffusion model has (conditionally) normally distributed jumps in equity returns. In this chapter, we find that an affine-jump diffusion model with double-exponential distributed jumps fits the data better.

Interestingly, there is a drawback of using normal distribution with a negative mean to model jump sizes, as in Merton’s model, because such a distribution does not have a monotone decreasing density for negative jumps. For example, if the jump mean is -3%, then with the normal distribution it is more likely to see a -2% negative jump than a -0.5% negative jump. This lack of monotonicity leads to a poor fitting for small jumps. That is also intuitively why Li, Wells and Yu (2008) find that Lévy type jumps fits the data better than the jump-diffusion model in Eraker, Johannes and Polson (2003) with normal jump sizes. However, with the double exponential distribution for jump sizes, the density is monotone decreasing for negative jumps, resulting in a better fit. In addition, the heavy-tail
feature of the double exponential distribution also helps generate large jumps during the crisis. For a more detailed discussion, see Section 3.2.1.

We found that a simple affine jump-diffusion model with both stochastic volatility and double-exponential-jump sizes in returns fits both S&P 500 and Nasdaq 100 daily return data well. In fact, the model outperforms existing ones (in particular models with Lévy jumps and affine jump-diffusion models with normal jump sizes) for both the Nasdaq 100 and S&P 500 returns during the Crisis, and is comparable for the S&P 500 returns before the Crisis; see Table 3.7 and Figures 3.4 and 3.5 for S&P 500 and Table 3.10 and Figures 3.8 and 3.9 for Nasdaq 100.

Therefore, our answer to the second question is that there seems to be finite number of large jumps in equity returns. In short, affine jump-diffusion model with a proper jump size distribution can fit equity return data well both before and during the crisis.

There is a large literature on jump risk of equity returns. Besides different objectives and answering different questions, this chapter is distinctive from the existing literature in terms of data, model specification, and econometric methodology. (1) For data selection, existing studies typically focus on the short episodes following the market turmoil from 1980’s to early 2000’s, say, the 1987 crash, the LTCM crisis, and the tech-bubble burst. Bates (2012) covers a period of 2007–2010, but in that subsample he does not explicitly report model fits or parameter estimates. This chapter uses prolonged data from 1980–2011, including the period after the outbreak of the financial crisis in 2007–2011, for which we report model fits as well as the changes of parameters before and during the crisis.
Moreover, existing work typically relies on options data or uses a combined data set of both equity returns and options. This chapter uses only returns data. As pointed out by Bates (2012), studies only on returns data are ‘of interest in its own right’ and are important in testing the compatibility of option prices with their underlying asset prices. (2) In terms of model specification, we estimated a new model, the Stochastic Volatility with Double Exponential Jumps (SV-DEJ) model, which has never been studied before except in Bates (2012). But since model complexity is not quantified and penalized in Bates (2012), models with more complex features such as time-varying intensity and autocorrelations are favored. To the contrary, in this chapter we uses the Deviance Information Criterion (DIC), which explicitly takes into account both model fit and model complexity in model selection; see Section 3.3.2. (3) In terms of econometric methods, we use the Bayesian MCMC inference, which is different from the MLE or GMM approaches in Bates (2012), Bakshi, Cao and Chen (1997), Chernov, Gallant, Ghysels and Tauchen (2003), Pan (2002), among others. Bayesian MCMC imputes latent variables of the models, such as jump times and jump sizes, which is beneficial to our analysis. For imputed jumps, see Figures 6 and 7.

The rest of the chapter is organized as follows. In Section 3.2, we first provide intuition of why double-exponentially distributed jump sizes may fit the data better, and then formally introduce models that have been studied in the literature and propose the new SV-DEJ model. In Section 3.3, we outline the Bayesian MCMC inference procedure utilized in our study and discuss model diagnostics to evaluate and compare model performance. Section
3.4 reports empirical results for both the S&P 500 index returns and the Nasdaq 100 index returns. Section 3.5 concludes. The appendix provides details of MCMC implementation.

### 3.2 The Models

#### 3.2.1 Intuition

Our empirical studies in Section 3.4 show that an affine jump-diffusion model with double-exponentially distributed jump sizes outperforms a model with normal jumps in equity returns both before and during the crisis. Intuition behind this lies in the differences between the two jump-size distributions.

1. In terms of small jumps, the normal distribution does not have monotone structure. For example, it can be seen from Figure 3.1 that, if jump sizes are normally distributed with mean $-2\%$ (as in the Merton’s jump model studied in previous literature), jumps down $-2\%$ is more likely to occur (i.e. has a higher density) than jumps down just $-1\%$.

2. To the contrary, with double-exponential distribution, larger negative jumps are always less likely to occur, disregard of the values of parameters. The monotone structure of double-exponential jumps then provides better model fit of small jumps in equity returns. That is an intuitive reason that Li, Wells and Yu (2008) find that Lévy-type jumps fits the data better than the jump-diffusion model in Eraker, Johannes and Polson (2003) with normal jump with negative means, and the model with double-exponential jumps have similar performance before the crisis as the model with Lévy-type jumps (see Section 3.4).
Figure 3.1: The Lack of Monotonicity for Normal Densities.

The following figures show that normal distribution does not have monotone structure, especially for small negative values. Figure (a) is a normal density with mean $-2$ and standard deviation $1.5$. Figure (b) is a double exponential density $f(x) = (p/\eta^+)e^{-x/\eta^+}1(x > 0) + ((1-p)/\eta^-)e^{x/\eta^-}1(x \leq 0)$ with $p = 1/3$, $\eta^+ = 1$ and $\eta^- = 2$. The negative parts of the densities are highlighted.

(2) In terms of large jumps, the double-exponential distribution is also suitable because it has heavy tails. A good intuition may be obtained by simply comparing the quantiles for both standardized Laplace (with a symmetric density of exponential-type tails $f(x) = (1/2)e^{-x}1(x > 0) + (1/2)e^{x}1(x < 0)$) and standardized t distributions (with power-type tail and typically considered to be heavy) with the same mean and same variance. The right quantiles for the Laplace and normalized t densities with degrees of freedom from 3 to 7 are given in Table 3.1 (taken from Heyde and Kou, 2004), which show the heavy-tail feature of the exponential distribution.

In light of the above, Heyde and Kou (2004) argue that it is difficult to distinguish exponential-type tails from power-type tails from empirical data unless one has extremely large sample size, perhaps in the order of tens of thousands or even hundreds of thousands.
Table 3.1: The (right) quantiles for the Laplace and normalized t densities.

Far at the right-tail, say, at the probability of 0.01%, the quantile of the Laplace density is larger than the quantile of t7 density, even when theoretically the Student-t density should have a asymptotically heavier tail than the exponential-type distribution.

<table>
<thead>
<tr>
<th>Probability</th>
<th>Laplace</th>
<th>t7</th>
<th>t6</th>
<th>t5</th>
<th>t4</th>
<th>t3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1%</td>
<td>2.77</td>
<td>2.53</td>
<td>2.57</td>
<td>2.61</td>
<td>2.65</td>
<td>2.62</td>
</tr>
<tr>
<td>0.1%</td>
<td>4.39</td>
<td>4.04</td>
<td>4.25</td>
<td>4.57</td>
<td>5.07</td>
<td>5.90</td>
</tr>
<tr>
<td>0.01%</td>
<td>6.02</td>
<td>5.97</td>
<td>6.55</td>
<td>7.50</td>
<td>9.22</td>
<td>12.82</td>
</tr>
<tr>
<td>0.001%</td>
<td>7.65</td>
<td>8.54</td>
<td>9.82</td>
<td>12.04</td>
<td>16.50</td>
<td>27.67</td>
</tr>
</tbody>
</table>

The heavy-tail feature of the double exponential distribution helps to fit the large jumps during the crisis, leading to a better performance than that of the model with Lévy-type jumps in Li, Wells and Yu (2008).

### 3.2.2 Model Formulation

A general model for asset returns $y_t$ in discrete time is

$$
\begin{align*}
\begin{cases}
  y_{t+1} &= y_t + \mu \Delta + \sqrt{v_t \Delta} \epsilon^y_{t+1} + J^y_{t+1} \\
v_{t+1} &= v_t + \kappa (\theta - v_t) \Delta + \sigma v \sqrt{v_t \Delta} \epsilon^v_{t+1} + J^v_{t+1}
\end{cases}
\end{align*}
$$

(3.1)

where $\Delta$ is the time length, $v_t$ stochastic volatility, $\epsilon^y_{t+1}$ and $\epsilon^v_{t+1}$ diffusion noises with normal distribution such that $\text{corr}(\epsilon^y_{t+1}, \epsilon^v_{t+1}) = \rho$. The jump part $(J^y_{t+1}, J^v_{t+1})$ is independent of the diffusion part $(\epsilon^y_{t+1}, \epsilon^v_{t+1})$. Various specifications of $J^y_{t+1}$ and $J^v_{t+1}$ lead to different models.

A. **Finite number of moderate jumps:** Stochastic Volatility with Correlated Merton’s
Jumps (SV-MJ-JV). In the model, $J_{t+1} = \xi_{t+1} N_{t+1}$, $J_{t+1} = \xi_{t+1} N_{t+1}$, where $P(N_{t+1} = 1) = 1 - P(N_{t+1} = 0) = \lambda \Delta$, $\xi_{t+1} \sim \exp(\mu_v)$, $\xi_{t+1} \sim N(\mu_v + \rho v \xi_{t+1}, \sigma_v^2)$. This model is studied in Eraker, Johannes and Polson (2003). Jump-size distribution is moderate in the sense that it is conditionally normal, a probability distribution without heavy-tails.

**B. Infinite number of small jumps:** Stochastic Volatility with Variance Gamma Jumps (SV-VG). In this model, jumps in return follow a discrete variance gamma process: $J_{t+1} = \gamma G_{t+1} + \sigma \sqrt{G_{t+1}} \epsilon_{t+1}$, where $\epsilon_{t+1} \sim N(0, 1)$ and $G_{t+1} \sim \Gamma(\Delta, \nu)$. $\epsilon_{t+1}$ and $G_{t+1}$ are independent of each other; there is no jumps in volatility, i.e. $J_{t+1} = 0$. The continuous variance gamma process is an infinite-activity Levy process, and has infinite number of small jumps in any given finite-time horizon. Li, Wells and Yu (2008) find that SV-VG fits the S&P 500 returns from 1980 to 2000 better than SV-MJ-JV.

The two models above only consider small or moderate jump sizes in returns. In this chapter, we propose a new model of large jump sizes with monotone structure in returns. In particular, the jump sizes have a double-exponential distribution. We are interested to see whether we can fit the returns data better both before and during the financial crisis with this new model.

**C. Finite number of large and monotonic jumps in returns:** Stochastic Volatility with Double-Exponential Jumps (SV-DEJ). In this model, jumps sizes $J_{t+1}$ follow a double-exponential distribution. The notation $\sim \exp(m)$ means the distribution is exponential with mean $m$.

1Eraker, Johannes and Polson (2003) also consider SV-IMJ (where jumps in returns and volatility are governed by two independent Poisson processes) and find that both SV-MJ-JV and SV-IMJ lead to similar fit to the data of S&P 500 returns from 1980 to 1999. However, since SV-IMJ is harder to be estimated, Eraker, Johannes and Polson (2003) prefer SV-MJ-JV.

3The notation $\Gamma(\alpha, \beta)$ means the distribution is gamma with parameter $\alpha$, $\beta$ and density $\frac{1}{\beta^\alpha} e^{\alpha-1} e^x / \beta$. 

51
exponential distribution, i.e.

\[ J_{t+1}^y = 1_{(N_{t+1}=1)}\xi_{t+1}^{y+} + 1_{(N_{t+1}=-1)}(-\xi_{t+1}^{y-}), \]

\[ \xi_{t+1}^{y+} \sim \exp(\eta^+), \quad \xi_{t+1}^{y-} \sim \exp(\eta^-), \quad N_{t+1} = \begin{cases} 1, & \text{w.p. } \lambda^+ \Delta; \\ 0, & \text{w.p. } 1 - \lambda^+ \Delta - \lambda^- \Delta; \\ -1, & \text{w.p. } \lambda^- \Delta. \end{cases} \]

There is no jumps in volatility, i.e. \( J_{t+1}^v = 0 \).

An important feature of Model C, different from existing ones, is that the jump sizes have the double-exponential distribution. A simpler model without stochastic volatility is proposed by Kou (2002) with emphasis on option pricing for path-dependent options.

Note that both our proposed model and SV-MJ-JV are special cases of affine jump-diffusion models, but SV-VG is not. Affine jump-diffusion models provide analytical tractability, and their model parameters have natural economic interpretations. For example, in Model C, \( \lambda^+ \) and \( \lambda^- \) are the positive and negative jump rates, respectively, and \( \eta^+ \) and \( \eta^- \) are the means of positive and negative jumps, respectively.

Model C is a simple affine jump-diffusion model with constant jump intensity rates and without jumps in volatility, thus simple to be implemented. More sophisticated models might have jump rates depending on volatility (Eraker (2004), Pan (2002)), or jumps in volatility (Eraker (2004), Broadie, Chernov and Johannes (2007)). We did not add these complex features to Model C because of two reasons: (1) Surprisingly, even in a relatively simple form, Model C fits both the pre- and post-crisis data well; see Section 3.4. Thus, we
have less motivation to incorporate more complicated features to Model C. (2) The focus of the chapter is on jumps in equity returns, not on volatility structures (such as jumps in volatility, the impact of volatility on jump rates, etc.). To fully investigate volatility structures, one perhaps needs to use option data, as options are more informative about volatilities. Indeed, all the papers aforementioned use option data. Following Bates (2012), we consider it deserves a separate paper to study volatility structures before and during the crisis using both equity and option data.

3.3 Econometric Methodology

3.3.1 Bayesian MCMC Inference

To perform statistical inference on the models above, we use the Bayesian Monte Carlo Markov Chain (MCMC) method; see a survey by Johannes and Polson (2003). In MCMC, statistical inference is performed via samples from the posterior distribution, which are obtained by constructing a Markov chain \( X(m) \), where \( m \) denotes the number of iterations, such that \( X(m)|X(0) \rightarrow X(\infty) \sim F \), as \( m \) goes to infinity, where \( F \) is the posterior distribution and independent of \( X(0) \).

In practice, we choose a large number \( m \) and use \( X(m), X(m + 1), \ldots, X(m + k) \) as approximated i.i.d. samples from \( F \). In order to ensure that the results are not sensitive to initial value \( X(0) \), we run multiple (e.g. 100) Markov chains with different starting points \( X(0) \).
Table 3.2: Conditional Posteriors, prior means and standard deviations for SV-DEJ.

(1) The parameters $\eta^+, \eta^-, \lambda^+$ and $\lambda^-$ are unique to the SV-DEJ model. For the details of the conditional posterior distributions (such as the formulae for hyper-parameters) as well as rigorous proofs, see Appendix B. (2) The parameters $\mu$, $\theta$, $\kappa$, $\sigma_v$ and $\rho$ are the parameters of the SV-DEJ model that are shared with SV-MJ-JV. Proofs of these conditional posteriors are provided in Li, Wells and Yu (2008) and Eraker, Johannes and Polson (2003); hence in Appendix B the details of these conditional posteriors are listed without proofs.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Posterior Distribution</th>
<th>Prior Mean</th>
<th>Prior Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\eta^+$</td>
<td>(conjugate) Inverse Gamma</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>$\eta^-$</td>
<td>(conjugate) Inverse Gamma</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>($\lambda^+, \lambda^-$)</td>
<td>(conjugate) Dirichlet (0.0455, 0.0455)</td>
<td>(0.0311, 0.0311)</td>
<td></td>
</tr>
<tr>
<td>$\mu$</td>
<td>(conjugate) Normal</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>$\theta$</td>
<td>(conjugate) Truncated Normal</td>
<td>0.7979</td>
<td>0.6028</td>
</tr>
<tr>
<td>$\kappa$</td>
<td>(conjugate) Truncated Normal</td>
<td>0.7979</td>
<td>0.6028</td>
</tr>
<tr>
<td>($\sigma_v, \rho$)</td>
<td>$\phi_v</td>
<td>w_v$ Normal, $w_v$ Inverse Gamma, $\sqrt{w_v}/2\infty$</td>
<td>0</td>
</tr>
</tbody>
</table>

Bayesian MCMC imputes latent variables of the model, such as jump times and jump sizes, which is beneficial to our analysis. For the same reason, Bayesian MCMC is applied in Eraker, Johannes and Polson (2003) and Li, Wells and Yu (2008) to provide inference for the SV-MJ-JV and SV-VG models. Therefore, we present only the MCMC procedure for the SV-DEJ model. Table 3.2 lists the conditional posteriors for the model parameters of SV-DEJ, of which priors are mostly conjugate. For other parameters we use the same priors as in Eraker, Johannes and Polson (2003) and Li, Wells and Yu (2008). For the detail calculation of the posteriors, refer to Appendix B.

To check the validity of the MCMC method, we conduct a simulation study using the data of 5000 observations (i.e. roughly 20 years worth of daily data); details can be found
Table 3.3: Simulation study of Bayesian MCMC

In the simulation study we use one simulated dataset of 5000 observations (i.e. roughly 20 years worth of daily data). On the same data set with the same priors, we run 100 Markov chains (each with a different starting point) each with 50000 iterations. For each Markov chain, the first 49900 were discarded as burn in. The resulting \(100 \times 100 = 10000\) samples of parameters in total from the 100 chains are considered 10000 random draws from the posterior distribution, from which we obtain the estimates of the posterior mean and standard deviation. We also obtain 100 sets of model residues from the 100 Markov chains, from which we perform the KS test and report its average p-values.

Simulation Results on the MCMC estimators

**Panel A: SV-MJ-JV**

<table>
<thead>
<tr>
<th>(\mu)</th>
<th>(\kappa)</th>
<th>(\theta)</th>
<th>(\sigma_v)</th>
<th>(\rho)</th>
<th>(\mu_y)</th>
<th>(\sigma_y)</th>
<th>(\lambda)</th>
<th>(\mu_v)</th>
<th>(\rho_J)</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>0.05</td>
<td>0.015</td>
<td>0.81</td>
<td>0.1</td>
<td>-0.4</td>
<td>-3.0</td>
<td>3.5</td>
<td>0.015</td>
<td>1.0</td>
</tr>
<tr>
<td>Posterior Mean</td>
<td>0.0503</td>
<td>0.0162</td>
<td>0.9677</td>
<td>0.1071</td>
<td>-0.3647</td>
<td>3.0923</td>
<td>3.0968</td>
<td>0.0146</td>
<td>1.0527</td>
</tr>
<tr>
<td>Posterior S.D.</td>
<td>0.0065</td>
<td>0.0027</td>
<td>0.0528</td>
<td>0.0032</td>
<td>0.0126</td>
<td>0.9038</td>
<td>0.2940</td>
<td>0.0038</td>
<td>0.1556</td>
</tr>
</tbody>
</table>

Percentage of Rejection of the KS test: 9%, Average p-value: 0.4587

**Panel B: SV-VG**

<table>
<thead>
<tr>
<th>(\mu)</th>
<th>(\kappa)</th>
<th>(\theta)</th>
<th>(\sigma_v)</th>
<th>(\rho)</th>
<th>(\gamma)</th>
<th>(\sigma)</th>
<th>(\nu)</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>0.05</td>
<td>0.015</td>
<td>0.81</td>
<td>0.1</td>
<td>-0.4</td>
<td>-0.01</td>
<td>0.4</td>
</tr>
<tr>
<td>Posterior Mean</td>
<td>0.0551</td>
<td>0.0169</td>
<td>0.7949</td>
<td>0.1012</td>
<td>-0.3920</td>
<td>-0.0103</td>
<td>0.4075</td>
</tr>
<tr>
<td>Posterior S.D.</td>
<td>0.0181</td>
<td>0.0025</td>
<td>0.0346</td>
<td>0.0068</td>
<td>0.0631</td>
<td>0.0098</td>
<td>0.0231</td>
</tr>
</tbody>
</table>

Percentage of Rejection of the KS test: 19%, Average p-value: 0.2239

**Panel C: SV-DEJ**

<table>
<thead>
<tr>
<th>(\mu)</th>
<th>(\kappa)</th>
<th>(\theta)</th>
<th>(\sigma_v)</th>
<th>(\rho)</th>
<th>(\eta^+)</th>
<th>(\eta^-)</th>
<th>(\lambda^+)</th>
<th>(\lambda^-)</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>0.05</td>
<td>0.015</td>
<td>0.81</td>
<td>0.1</td>
<td>-0.4</td>
<td>4.0</td>
<td>4.6</td>
<td>0.004</td>
</tr>
<tr>
<td>Posterior Mean</td>
<td>0.0512</td>
<td>0.0158</td>
<td>0.8211</td>
<td>0.0964</td>
<td>-0.3819</td>
<td>4.1254</td>
<td>4.4898</td>
<td>0.0042</td>
</tr>
<tr>
<td>Posterior S.D.</td>
<td>0.0084</td>
<td>0.0025</td>
<td>0.0127</td>
<td>0.0057</td>
<td>0.0214</td>
<td>0.1602</td>
<td>0.1683</td>
<td>0.0056</td>
</tr>
</tbody>
</table>

Percentage of Rejection of the KS test: 11%, Average p-value: 0.3755

in the caption of Table 3.3. The estimators for all the models using the priors in Table 3.2 are reported in Table 3.3, which suggests that the MCMC inference performs well.

### 3.3.2 Model Diagnostics

To check the goodness-of-fit of the models, we use the following diagnostic methods.

1. The Kolmogorov-Smirnov (KS) test for normality of normalized residues

\[
e_{t+1}^y = \frac{y_{t+1} - y_t - \mu \Delta - J_{t+1}^y}{\sqrt{v_t \Delta}}.
\]
This is the same method used in Li, Wells and Yu (2008). If a model in question has good fit to the data, then normalized residues $\epsilon_{t+1}$ from that model should follow a standard normal distribution. In the simulation study, for each of the models one KS test is performed on the residues $\epsilon_{t+1}$ obtained from each of the 100 simulation runs, resulting in 100 rejections/fails to reject from which we obtain the percentage of rejection and the average p-value. A lower percentage of rejection and a higher average p-value thus suggest a better goodness-of-fit of the model. The percentage of rejection and average p-values of the KS tests for the simulation study are also reported in Table 3.3.

(2) The Deviance Information Criterion (DIC). There are three conventional criteria for model selection: the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC) and the Bayes Factor (BF). In the models above, the number of latent variables (volatilities and jumps) grows with the number of observations, precluding the use of the AIC and the BIC (see Berg, Meryer, Yu (2004)). Moreover, calculating the BF may be complicated since Bayes factors require the integration over all unknown model parameters and latent variables (three $T \times 1$-dimension variables in each of the three models).

In this chapter, we adopt a simple Bayesian model selection criteria, the Deviance Information Criterion (DIC), recently proposed by Spiegelhalter, Best, Carlin and van der Linde (2002) for model comparison; asymptotic properties of DIC are studied in Ando (2007, 2012). Berg, Meryer, Yu (2004) use DIC to study stochastic volatility models. Two advantages of DIC are particularly relevant in our study. (i) The DIC is easy to interpret. By definition (see (3.2) below), DIC takes into account both model fit and model complexity.
In particular, the effective number of model parameter $p_D$ (see (3.3)) includes a penalizing term for model complexity, generalizing the AIC and relating to the BIC. (ii) The DIC is straightforward to calculate. Given posterior samples from MCMC, quantities required for the DIC are ‘almost trivial to compute’ (see the comment in Spiegelhalter, Best, Carlin and van der Linde (2002) and also the formula in (3.4)).

Formally, the DIC is defined as

$$DIC := E[-2 \log l(Y_{0:T})] + p_D, \quad (3.2)$$

where $T$ is the number of observations, $l(Y_{0:T})$ the likelihood of data $Y_{0:T} = (y_0, \ldots, y_T)$ given model parameters $\Theta$ and latent volatilities $V_{0:T} = (v_0, \ldots, v_T)$ and latent jumps $J_{1:T}^y = (J_1^y, \ldots, J_T^y)$. The effective number of parameters is defined as

$$p_D := -E[2 \log l(Y_{0:T})] + 2 \log p(Y_{0:T}|\bar{V}_{0:T}, \bar{J}_{1:T}^y, \bar{\Theta}), \quad (3.3)$$

where $p(Y_{0:T}|\bar{V}_{0:T}, \bar{J}_{1:T}^y, \bar{\Theta})$ is the likelihood given the corresponding posterior means $\bar{V}_{0:T}$, $\bar{J}_{1:T}^y$, $\bar{\Theta}$. In the above, both expectations are taken over posterior distribution

$$V_{0:T}, J_{1:T}^y, \Theta|Y_{0:T}.$$ 

To approximate the expectation, we use the sample averages of posterior distributions from
Table 3.4: Deviance Information Criteria (DIC) on Model Fitting of Simulated Data

DIC is smallest when the true model is used. In the simulation study, a sample path with length $T = 5000$ is simulated using the SV-MJ-JV model. We then perform Bayesian MCMC of all of the three models on the sample path and obtain three DICs by Equation (3.2). The same is repeated for sample paths simulated by the SV-VG and the SV-DEJ models, resulting in a total of nine DICs.

<table>
<thead>
<tr>
<th>Model Fitted</th>
<th>True Model</th>
<th>SV-MJ-JV</th>
<th>SV-VG</th>
<th>SV-DEJ</th>
</tr>
</thead>
<tbody>
<tr>
<td>SV-MJ-JV</td>
<td>14706</td>
<td>13990</td>
<td>14680</td>
<td></td>
</tr>
<tr>
<td>SV-VG</td>
<td>14729</td>
<td>13644</td>
<td>14957</td>
<td></td>
</tr>
<tr>
<td>SV-DEJ</td>
<td>14813</td>
<td>13719</td>
<td>14468</td>
<td></td>
</tr>
</tbody>
</table>

MCMC (see Spiegelhalter, Best, Carlin and van der Linde (2002)).

\[
DIC = -4E[\log l(Y_{0:T})] + 2\log p(Y_{0:T}|\bar{V}_{0:T}, \bar{J}_{1:T}, \bar{\Theta})
\approx -4 \left[ \frac{1}{M} \sum_{m=1}^{M} \sum_{t=0}^{T-1} \left( -\frac{1}{2} \log(2\pi \Delta \bar{v}_t^{(m)}) - \frac{(y_{t+1} - y_t - \mu_{t+1}^{(m)} - (J_{t+1}^{y})^{(m)})^2}{2\Delta \bar{v}_t^{(m)}} \right) \right]
+ 2 \sum_{t=0}^{T-1} \left( -\frac{1}{2} \log(2\pi \Delta \bar{v}_t) - \frac{(y_{t+1} - y_t - \bar{\mu} - \bar{J}_{t+1})^2}{2\Delta \bar{v}_t} \right) \tag{3.4}
\]

Note that in calculating the likelihood $l(Y_{0:T})$, we use the fact that $\epsilon_{t+1}^{y}$, $t = 0, \ldots, T - 1$ are independent and normally distributed in all three models (see (3.1)). A smaller value of DIC indicates better model fit.

In the simulation study, a sample path of $T = 5000$ is generated using the SV-MJ-JV model. We then fit each of the three models via Bayesian MCMC on the sample path and obtain three DICs. The same is repeated for sample paths generated by the SV-VG and the SV-DEJ models, resulting in a total of nine DICs. If DIC is a good diagnostics, its value should be the smallest when the true model is fitted. In each of the columns in Table 3.4, it is indeed the case.
(3) To further check normality of model residues, we compare the histograms of normalized model residues $e_{t+1}^y$ with the standard normal density. The histograms of the simulation study are reported in Figure 3.2.

Figure 3.2: Histograms of Normalized Model Residue in Simulation Study
Under each of three models, we compare the histograms of normalized model residues with standard normal density (the solid curve). The three figures represent the SV-MJ-JV model (top left), the SV-VG model (top right) and the SV-DEJ model (bottom), respectively.

(4) We examine QQ plots of actual data versus model-simulated $Y_{0:T}$ using estimated model parameters. QQ plots close to the straight line suggest a good model fit. In the simulation study, we compare the quantiles of $Y_{0:T}$ simulated using the true parameter
values versus quantiles of model-simulated $Y_{0:T}$ using the point estimators in Table 3.3.

The QQ plots of the simulation study are reported in Figure 3.3.

**Figure 3.3: QQ Plots of Simulation Study**

Under each of three models, we compare quantiles of sample paths simulated using the true parameter values with quantiles of sample paths simulated using the point estimated in Table 3.3. The three figures represent the SV-MJ-JV model (top left), the SV-VG model (top right) and the SV-DEJ model (bottom), respectively. In all of the three figures, the horizontal axis represents quantiles of returns simulated with true parameters, while the vertical axis represents quantiles of returns simulated with estimated parameters.
3.4 Empirical Results

3.4.1 The Data

The first data set used in this chapter consists of the daily returns of S&P 500 from January 1980 to December 2011. To better relate to Li, Wells and Yu (2008) and Eraker, Johannes and Polson (2003), for S&P 500 we consider three separate periods: (i) 1980.01–2000.12, which coincides exactly with the data set considered in Li, Wells and Yu (2008), and roughly with Eraker, Johannes and Polson (2003), who consider 1980–1999; (ii) 2001.01–2007.07 (before the financial crisis); and (iii) 2007.08–2011.12 (during the financial crisis).

To check the robustness of our findings, we also use a second data set of daily returns of Nasdaq 100 from 2001.01 to 2011.12. To match the study of S&P 500, we consider two separate periods of Nasdaq 100: 2001.01–2007.07 and 2007.08–2011.12. Table 3.5 provides summary statistics for the daily (log) returns.\(^4\)

3.4.2 S&P 500

The point estimates of the model parameters, in particular, the positive and negative jump rates, and average positive and negative jump sizes before and during the crisis, are reported Table 3.6. For the period 1980 to 2000, the results match those in Li, Wells and

\(^4\)Following the convention in Li, Wells and Yu (2008), we multiply the daily returns by 100.
Table 3.5: Summary statistics of the data sets

<table>
<thead>
<tr>
<th>Data Periods</th>
<th>Mean</th>
<th>Volatility</th>
<th>Skewness</th>
<th>Kurtosis</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panal A: S&amp;P 500</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1980.01–2000.12</td>
<td>0.0476</td>
<td>1.0435</td>
<td>-2.3584</td>
<td>55.6080</td>
<td>-22.8997</td>
<td>8.7089</td>
</tr>
<tr>
<td>2001.01–2007.07</td>
<td>0.0076</td>
<td>1.0531</td>
<td>0.1159</td>
<td>5.9645</td>
<td>-5.0468</td>
<td>5.5744</td>
</tr>
<tr>
<td>Panal B: Nasdaq 100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2001.01–2007.07</td>
<td>-0.0206</td>
<td>1.9063</td>
<td>0.0976</td>
<td>6.2757</td>
<td>-8.6139</td>
<td>10.2727</td>
</tr>
<tr>
<td>2007.08–2011.12</td>
<td>0.0135</td>
<td>1.8055</td>
<td>-0.0984</td>
<td>8.3572</td>
<td>-11.1149</td>
<td>11.8493</td>
</tr>
</tbody>
</table>

Yu (2008) closely. In particular, we find significant leverage effect in all three models ($\rho$ between $-0.4$ and $-0.9$), consistent with previous studies (e.g. Duffee (1995)).

The results of KS test and DICs are summarized in Table 3.7. The comparison between histograms of normalized model residues $e_{t+1}$ and the standard normal density can be found in Figure 3.4. The QQ plots for residuals are given in Figure 3.5.

We summarize the empirical findings as follows.

(i) Judging from the KS test and DICs, the SV-DEJ model outperforms all other models during the financial crisis. (ii) The QQ plots seem to confirm that the SV-DEJ model provides the best tail fit. In particular, better fit in the tails by the SV-DEJ model becomes apparent for the post-crisis data, suggesting that the better performance of the SV-DEJ model comes from capturing large jumps. (iii) During normal time period and before the crisis, from 1980 to 2007, the SV-DEJ model and the SV-VG model seem to perform equally well by and large, with SV-VG slightly better, which is consistent with the results in Bates (2012) for the CRSP value-weighted returns up to 2006. Hence it appears that the
Table 3.6: S&P 500: Model Parameters Estimates

We use 100 Markov chains with different starting points. For each of the models, we estimate the posterior mean and posterior standard deviation using the last 100 random draws from each of the Markov chains (each Markov chain has 50000 iteration in total, while we discard the first 49900 as burn-in).

**Panel A: 1980.01–2000.12**

<table>
<thead>
<tr>
<th>Model</th>
<th>Posterior Mean</th>
<th>Posterior S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SV-MJ-JV</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Posterior Mean</td>
<td>0.0503</td>
<td>0.0053</td>
</tr>
<tr>
<td></td>
<td>0.0277</td>
<td>0.0030</td>
</tr>
<tr>
<td></td>
<td>0.6697</td>
<td>0.0302</td>
</tr>
<tr>
<td></td>
<td>0.1032</td>
<td>0.0029</td>
</tr>
<tr>
<td></td>
<td>-0.4830</td>
<td>0.0158</td>
</tr>
<tr>
<td></td>
<td>-1.4946</td>
<td>0.7107</td>
</tr>
<tr>
<td></td>
<td>2.1938</td>
<td>0.2711</td>
</tr>
<tr>
<td></td>
<td>0.0067</td>
<td>0.0034</td>
</tr>
<tr>
<td></td>
<td>1.0192</td>
<td>0.1215</td>
</tr>
<tr>
<td></td>
<td>-0.8103</td>
<td>0.7930</td>
</tr>
<tr>
<td><strong>SV-VG</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Posterior Mean</td>
<td>0.0755</td>
<td>0.0150</td>
</tr>
<tr>
<td></td>
<td>0.0119</td>
<td>0.0024</td>
</tr>
<tr>
<td></td>
<td>0.8953</td>
<td>0.0437</td>
</tr>
<tr>
<td></td>
<td>0.1051</td>
<td>0.0095</td>
</tr>
<tr>
<td></td>
<td>-0.5687</td>
<td>0.0635</td>
</tr>
<tr>
<td></td>
<td>-0.0468</td>
<td>0.0096</td>
</tr>
<tr>
<td></td>
<td>0.3307</td>
<td>0.0203</td>
</tr>
<tr>
<td></td>
<td>5.9248</td>
<td>0.1826</td>
</tr>
<tr>
<td><strong>SV-DEJ</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Posterior Mean</td>
<td>0.0479</td>
<td>0.0071</td>
</tr>
<tr>
<td></td>
<td>0.0162</td>
<td>0.0100</td>
</tr>
<tr>
<td></td>
<td>0.9144</td>
<td>0.0123</td>
</tr>
<tr>
<td></td>
<td>0.1076</td>
<td>0.0032</td>
</tr>
<tr>
<td></td>
<td>-0.4527</td>
<td>0.0210</td>
</tr>
<tr>
<td></td>
<td>0.2690</td>
<td>0.1597</td>
</tr>
<tr>
<td></td>
<td>2.5991</td>
<td>0.1719</td>
</tr>
<tr>
<td></td>
<td>0.0257</td>
<td>0.0060</td>
</tr>
<tr>
<td></td>
<td>0.0090</td>
<td>0.0007</td>
</tr>
</tbody>
</table>

**Panel B: 2001.01–2007.07**

<table>
<thead>
<tr>
<th>Model</th>
<th>Posterior Mean</th>
<th>Posterior S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SV-MJ-JV</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Posterior Mean</td>
<td>0.0042</td>
<td>0.0023</td>
</tr>
<tr>
<td></td>
<td>0.0193</td>
<td>0.0033</td>
</tr>
<tr>
<td></td>
<td>1.0444</td>
<td>0.0462</td>
</tr>
<tr>
<td></td>
<td>0.1404</td>
<td>0.0073</td>
</tr>
<tr>
<td></td>
<td>-0.7553</td>
<td>0.0269</td>
</tr>
<tr>
<td></td>
<td>-1.3753</td>
<td>2.9737</td>
</tr>
<tr>
<td></td>
<td>2.0348</td>
<td>0.0817</td>
</tr>
<tr>
<td></td>
<td>0.0037</td>
<td>0.0012</td>
</tr>
<tr>
<td></td>
<td>1.0000</td>
<td>0.9528</td>
</tr>
<tr>
<td></td>
<td>0.0918</td>
<td>0.7758</td>
</tr>
<tr>
<td><strong>SV-VG</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Posterior Mean</td>
<td>0.0331</td>
<td>0.0118</td>
</tr>
<tr>
<td></td>
<td>0.0141</td>
<td>0.0038</td>
</tr>
<tr>
<td></td>
<td>1.4864</td>
<td>0.0623</td>
</tr>
<tr>
<td></td>
<td>0.1780</td>
<td>0.0138</td>
</tr>
<tr>
<td></td>
<td>-0.8617</td>
<td>0.0294</td>
</tr>
<tr>
<td></td>
<td>-0.0389</td>
<td>0.0085</td>
</tr>
<tr>
<td></td>
<td>0.1875</td>
<td>0.0127</td>
</tr>
<tr>
<td></td>
<td>4.8425</td>
<td>0.0866</td>
</tr>
<tr>
<td><strong>SV-DEJ</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Posterior Mean</td>
<td>0.0269</td>
<td>0.0021</td>
</tr>
<tr>
<td></td>
<td>0.0176</td>
<td>0.0010</td>
</tr>
<tr>
<td></td>
<td>0.9734</td>
<td>0.0261</td>
</tr>
<tr>
<td></td>
<td>0.1336</td>
<td>0.0045</td>
</tr>
<tr>
<td></td>
<td>-0.7208</td>
<td>0.0134</td>
</tr>
<tr>
<td></td>
<td>0.2460</td>
<td>0.0635</td>
</tr>
<tr>
<td></td>
<td>0.4121</td>
<td>0.0522</td>
</tr>
<tr>
<td></td>
<td>0.0275</td>
<td>0.0027</td>
</tr>
<tr>
<td></td>
<td>0.0572</td>
<td>0.0041</td>
</tr>
</tbody>
</table>

**Panel C: 2007.08–2011.12**

<table>
<thead>
<tr>
<th>Model</th>
<th>Posterior Mean</th>
<th>Posterior S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SV-MJ-JV</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Posterior Mean</td>
<td>0.0738</td>
<td>0.0037</td>
</tr>
<tr>
<td></td>
<td>0.0234</td>
<td>0.0026</td>
</tr>
<tr>
<td></td>
<td>1.4980</td>
<td>0.1490</td>
</tr>
<tr>
<td></td>
<td>0.2305</td>
<td>0.0180</td>
</tr>
<tr>
<td></td>
<td>-0.7388</td>
<td>0.0272</td>
</tr>
<tr>
<td></td>
<td>-1.6592</td>
<td>1.9092</td>
</tr>
<tr>
<td></td>
<td>1.9578</td>
<td>0.0790</td>
</tr>
<tr>
<td></td>
<td>0.0135</td>
<td>0.0029</td>
</tr>
<tr>
<td></td>
<td>1.1211</td>
<td>0.2444</td>
</tr>
<tr>
<td></td>
<td>-0.1188</td>
<td>0.5712</td>
</tr>
<tr>
<td><strong>SV-VG</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Posterior Mean</td>
<td>0.0655</td>
<td>0.0145</td>
</tr>
<tr>
<td></td>
<td>0.0292</td>
<td>0.0035</td>
</tr>
<tr>
<td></td>
<td>2.5191</td>
<td>0.0992</td>
</tr>
<tr>
<td></td>
<td>0.4042</td>
<td>0.0287</td>
</tr>
<tr>
<td></td>
<td>-0.7954</td>
<td>0.0353</td>
</tr>
<tr>
<td></td>
<td>-0.0076</td>
<td>0.0118</td>
</tr>
<tr>
<td></td>
<td>0.2046</td>
<td>0.0085</td>
</tr>
<tr>
<td></td>
<td>5.9176</td>
<td>0.1124</td>
</tr>
<tr>
<td><strong>SV-DEJ</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Posterior Mean</td>
<td>0.0602</td>
<td>0.0089</td>
</tr>
<tr>
<td></td>
<td>0.0243</td>
<td>0.0019</td>
</tr>
<tr>
<td></td>
<td>2.1473</td>
<td>0.0652</td>
</tr>
<tr>
<td></td>
<td>0.2722</td>
<td>0.0140</td>
</tr>
<tr>
<td></td>
<td>-0.7283</td>
<td>0.0295</td>
</tr>
<tr>
<td></td>
<td>0.2464</td>
<td>0.0485</td>
</tr>
<tr>
<td></td>
<td>0.5183</td>
<td>0.0933</td>
</tr>
<tr>
<td></td>
<td>0.0326</td>
<td>0.0023</td>
</tr>
<tr>
<td></td>
<td>0.0759</td>
<td>0.0063</td>
</tr>
</tbody>
</table>
Figure 3.4: S&P 500: Normalized Model Residues vs Standard Normal

It seems that, in particular during the financial crisis, the SV-DEJ model is better in terms of normality of normalized model residues, especially near zero. In the following graphs, the solid curves represents standard normal density. Data used is returns of S&P 500 Index, 1980.01-2000.12 for the left three graphs and 2007.08–2011.12 for the right three graphs. From top to bottom, models represented in the graphs are the SV-MJ-JV model, the SV-VG model and the SV-DEJ model.

Left: Before Crisis  
Right: During Crisis
Figure 3.5: S&P 500: QQ plots Before and During the Crisis

It appears that the SV-DEJ model provides best fit in both tails among models considered both before and during the financial crisis. In what follows are QQ plots between a sample path simulated from the corresponding model using point estimates in Table 3.6 and actual data (returns of S&P 500 Index, 1980.01-2000.12 for the left three graphs and 2007.08–2011.12 for the right three graphs). From top to bottom, models represented in the graphs are the SV-MJ-JV model, the SV-VG model and the SV-DEJ model. In all of the graphs above, quantiles are compared in the range $-4 \leq Z \leq 4$, where $Z$ is a random variable representing the actual returns.
monotonic structure of double-exponential jumps can also provide good fit to small jumps in the data, as prescribed by Lévy-jump models.

Based on the SV-DEJ model, the best fitted model, we report the approximated credible intervals of changes in jump parameters in Table 3.8. The left and right endpoints of 95% credible intervals for percentage of change are approximated by their simulation values. That is, we report the 2.5% and 97.5% quantiles in the pool of random draws of $\frac{\theta_2^{(m)}}{\theta_1^{(m)}}$ from the posterior distribution, where $\theta_2^{(m)}$ and $\theta_1^{(m)}$ denote the value of the $m$-th draw of the corresponding parameter from the posterior distribution by MCMC, before and during the crisis, respectively.

Given the 95% credible intervals, it appears that both positive and negative jump rates increased significantly during the crisis, while there is a mixed picture for the changes in average jump sizes before and during the crisis. To obtain a visual impression, we also plot the jumps imputed by the SV-DEJ model; see Figure 3.6.

3.4.3 Nasdaq 100

Parallel to the study of S&P 500, we perform the same analysis on the returns of Nasdaq 100. Point estimates of model parameters are reported in Table 3.9, while the results for KS test and DIC in Table 3.10. Histograms of normalized model residues can be found in Table 3.8. QQ plots are reported in Figure 3.9. Credible intervals for percentage of changes of jump parameters can be found in Table 3.11. Empirical results for Nasdaq 100 returns seem to be consistent with those for S&P 500.
Table 3.7: S&P 500: Kolmogorov-Smirnov Tests and Deviance Information Criteria
Both tests seem to favor the SV-DEJ model during the crisis. The SV-DEJ also performs reasonably well before the crisis. More specifically, the table reports the percentage of rejection of the Kolmogorov-Smirnov tests (KS) out of the 100 Markov chains, the average p-value of the KS tests and the Deviance Information Criterion (DIC) for each of the models for S&P 500 returns in each of the three periods considered. It can be seen that the SV-DEJ is with largest average p-value and lowest DIC during crisis (in boldface).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>SV-MJ-JV</td>
<td>81%</td>
<td>0.0034</td>
<td>15201.9</td>
<td></td>
</tr>
<tr>
<td>SV-VG</td>
<td>9%</td>
<td>0.4212</td>
<td>14874.4</td>
<td></td>
</tr>
<tr>
<td>SV-DEJ</td>
<td>15%</td>
<td><strong>0.4011</strong></td>
<td><strong>15066.5</strong></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panal B: 2001.01–2007.07</th>
<th>Model</th>
<th>Percentage of Rejection by KS</th>
<th>Average p-value of KS</th>
<th>DIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>SV-MJ-JV</td>
<td>78%</td>
<td>0.0035</td>
<td>4279.2</td>
<td></td>
</tr>
<tr>
<td>SV-VG</td>
<td>10%</td>
<td>0.4134</td>
<td>3814.0</td>
<td></td>
</tr>
<tr>
<td>SV-DEJ</td>
<td>19%</td>
<td><strong>0.3715</strong></td>
<td><strong>3908.2</strong></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>SV-MJ-JV</td>
<td>62%</td>
<td>0.0071</td>
<td>4106.3</td>
<td></td>
</tr>
<tr>
<td>SV-VG</td>
<td>96%</td>
<td>0.0001</td>
<td>4084.9</td>
<td></td>
</tr>
<tr>
<td>SV-DEJ</td>
<td>12%</td>
<td><strong>0.4125</strong></td>
<td><strong>4030.5</strong></td>
<td></td>
</tr>
</tbody>
</table>

Table 3.8: S&P 500: Change of Jump-Related Parameters
It appears that changes of both positive and negative jump rates increased significantly, while on average jump sizes did not increase. More precisely, we report in this table changes of jump-related parameters estimated for SV-DEJ before and during the crisis (i.e. 2001.01–2007.07 and 2007.08–2011.12, respectively) for the S&P 500 returns. The left and right endpoints of 95% credible intervals for percentage of change are approximated by their simulation values. That is, we report the 2.5% and 97.5% quantiles in the pool of random draws of $\theta_2^{(m)}/\theta_1^{(m)}$ from the posterior distribution, where $\theta_2^{(m)}$ and $\theta_1^{(m)}$ denote the value of the $m$-th draw of the corresponding parameter before and during the crisis, respectively.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Before Crisis</th>
<th>During Crisis</th>
<th>95% C.I. of the Percentage Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive Jump Rate $\lambda^+$</td>
<td>0.0275(daily)</td>
<td>0.0326(daily)</td>
<td>[11.67%, 25.95%]</td>
</tr>
<tr>
<td>Negative Jump Rate $\lambda^-$</td>
<td>6.9300(annually)</td>
<td>8.2151(annually)</td>
<td>[18.94%, 52.76%]</td>
</tr>
<tr>
<td>Mean Positive Jump Size $\eta^+$</td>
<td>0.2460</td>
<td>0.2464</td>
<td>[-30.65%, 22.22%]</td>
</tr>
<tr>
<td>Mean Negative Jump Size $\eta^-$</td>
<td>0.4121</td>
<td>0.5183</td>
<td>[-33.41%, 48.94%]</td>
</tr>
</tbody>
</table>
Figure 3.6: S&P 500: Jumps
Jumps (i.e. \(1_{(N_{t+1}=1)}(\xi_{t+1}^+ - \xi_{t+1}^-)\)) imputed by the SV-DEJ model for S&P 500. The vertical dotted line indicates the outbreak of the financial crisis (August 2007).

Figure 3.7: Nasdaq 100: Jumps
Jumps (i.e. \(1_{(N_{t+1}=1)}(\xi_{t+1}^+ - \xi_{t+1}^-)\)) imputed by the SV-DEJ model for Nasdaq 100. The dotted vertical line indicates the outbreak of the financial crisis (August 2007).
Table 3.9: Nasdaq 100: Model Parameter Estimates

For each of the models, we use 100 Markov chains with different starting points. We estimate the posterior mean and posterior standard deviation using the last 100 random draws from each of the Markov chains (i.e. 50000 iteration in total, discarding first 49900 as burn-in).

<table>
<thead>
<tr>
<th></th>
<th>SV-MJ-JV Model</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>SV-VG Model</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(\mu)</td>
<td>(\kappa)</td>
<td>(\theta)</td>
<td>(\sigma_v)</td>
<td>(\rho)</td>
<td>(\mu_y)</td>
<td>(\lambda)</td>
<td>(\mu_v)</td>
<td>(\rho_j)</td>
<td></td>
</tr>
<tr>
<td><strong>Posterior Mean</strong></td>
<td>0.0400</td>
<td>0.0071</td>
<td>1.3712</td>
<td>0.1272</td>
<td>-0.5833</td>
<td>-2.3703</td>
<td>2.0382</td>
<td>0.0059</td>
<td>0.9985</td>
<td>0.2456</td>
</tr>
<tr>
<td><strong>Posterior S.D.</strong></td>
<td>0.0029</td>
<td>0.0009</td>
<td>0.0868</td>
<td>0.0098</td>
<td>0.0423</td>
<td>1.2082</td>
<td>0.0723</td>
<td>0.0012</td>
<td>0.1320</td>
<td>0.8084</td>
</tr>
</tbody>
</table>

|                | SV-MJ-JV Model |               |               |               |               |               | SV-VG Model |               |               |               |
|                | \(\mu\)  | \(\kappa\)  | \(\theta\)  | \(\sigma_v\) | \(\rho\)  | \(\gamma\)  | \(\sigma\)  | \(\nu\)  |               |               |
| **Posterior Mean** | -0.0524 | 0.0128 | 2.8839 | 0.2716 | -0.3323 | 0.1103 | 0.2956 | 3.2920 |               |               |
| **Posterior S.D.** | 0.0814 | 0.0025 | 0.2074 | 0.0216 | 0.0807 | 0.0777 | 0.0609 | 0.2144 |               |               |

|                | SV-MJ-JV Model |               |               |               |               |               | SV-VG Model |               |               |               |
|                | \(\mu\)  | \(\kappa\)  | \(\theta\)  | \(\sigma_v\) | \(\rho\)  | \(\eta^+\)  | \(\eta^-\)  | \(\lambda^+\)  | \(\lambda^-\)  |               |
| **Posterior Mean** | 0.0663 | 0.0067 | 1.5902 | 0.1295 | -0.5608 | 0.2468 | 0.5374 | 0.0361 | 0.0755 |               |
| **Posterior S.D.** | 0.0067 | 0.0008 | 0.0773 | 0.0085 | 0.0501 | 0.0575 | 0.1014 | 0.0027 | 0.0057 |               |

Panal A: 2001.01–2007.07

|                | SV-MJ-JV Model |               |               |               |               |               | SV-VG Model |               |               |               |
|                | \(\mu\)  | \(\kappa\)  | \(\theta\)  | \(\sigma_v\) | \(\rho\)  | \(\mu_y\)  | \(\lambda\) | \(\mu_v\)  | \(\rho_j\)  |               |
| **Posterior Mean** | 0.0908 | 0.0317 | 1.7869 | 0.2729 | -0.7674 | -1.4604 | 1.7621 | 0.0231 | 1.7915 | -0.2252 |
| **Posterior S.D.** | 0.0062 | 0.0035 | 0.2157 | 0.0225 | 0.0227 | 0.6199 | 0.0689 | 0.0041 | 0.6068 | 0.2235 |

|                | SV-MJ-JV Model |               |               |               |               |               | SV-VG Model |               |               |               |
|                | \(\mu\)  | \(\kappa\)  | \(\theta\)  | \(\sigma_v\) | \(\rho\)  | \(\gamma\)  | \(\sigma\)  | \(\nu\)  |               |               |
| **Posterior Mean** | 0.0836 | 0.0304 | 2.7776 | 0.3445 | -0.7462 | -0.0222 | 0.2066 | 3.1564 |               |               |
| **Posterior S.D.** | 0.0136 | 0.0036 | 0.0547 | 0.0181 | 0.0366 | 0.0140 | 0.0163 | 0.1588 |               |               |

|                | SV-MJ-JV Model |               |               |               |               |               | SV-VG Model |               |               |               |
|                | \(\mu\)  | \(\kappa\)  | \(\theta\)  | \(\sigma_v\) | \(\rho\)  | \(\eta^+\)  | \(\eta^-\)  | \(\lambda^+\)  | \(\lambda^-\)  |               |
| **Posterior Mean** | 0.0960 | 0.0302 | 2.3906 | 0.3034 | -0.7078 | 0.2513 | 0.8213 | 0.0438 | 0.0847 |               |
| **Posterior S.D.** | 0.0092 | 0.0030 | 0.2520 | 0.0198 | 0.0347 | 0.0556 | 0.6515 | 0.0035 | 0.0095 |               |

69
Table 3.10: Nasdaq 100: Kolmogorov-Smirnov Tests and Deviance Information Criterion

Both tests seem to favor the SV-DEJ model during the crisis, with smallest DIC and largest average p-value by KS test. The SV-DEJ also performs reasonably well compared to the SV-VG model before the crisis. More specifically, the table reports the percentage of rejection of the Kolmogorov-Smirnov tests (KS) out of the 100 Markov chains, the average p-value of the KS tests and the Deviance Information Criterion (DIC) for each of the models for Nasdaq 100 returns in each of the three data periods considered.

<table>
<thead>
<tr>
<th>Panel A: 2001.01–2007.07</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
</tr>
<tr>
<td>SV-MJ-JV</td>
</tr>
<tr>
<td>SV-VG</td>
</tr>
<tr>
<td>SV-DEJ</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
</tr>
<tr>
<td>SV-MJ-JV</td>
</tr>
<tr>
<td>SV-VG</td>
</tr>
<tr>
<td>SV-DEJ</td>
</tr>
</tbody>
</table>

Table 3.11: Nasdaq 100: Change of Jump-Related Parameters

It appears that changes of both positive and negative jump rates increased significantly, while average jump sizes did not seem to change. More precisely, in this table we report the changes of jump-related parameters estimated for SV-DEJ before and during the crisis (i.e. 2001.01–2007.07 and 2007.08–2011.12, respectively) for the Nasdaq 100 returns. The left and right endpoints of 95% credible intervals for percentage of change are approximated by their simulation values. That is, we report the 2.5% and 97.5% quantiles in the pool of random draws of \( \theta_2^{(m)} / \theta_1^{(m)} \) from the posterior distribution, where \( \theta_2^{(m)} \) and \( \theta_1^{(m)} \) denote the value of the \( m \)-th draw of the corresponding parameter before and during the crisis, respectively.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Before Crisis</th>
<th>During Crisis</th>
<th>95% C.I. of the Percentage Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive Jumps</td>
<td>0.0361(daily)</td>
<td>0.0438(daily)</td>
<td>[23.78%, 47.94%]</td>
</tr>
<tr>
<td>Rate ( \lambda^+ )</td>
<td>9.0972(annually)</td>
<td>11.0376(annually)</td>
<td></td>
</tr>
<tr>
<td>Negative Jumps</td>
<td>0.0755(daily)</td>
<td>0.0847(daily)</td>
<td>[9.41%, 21.84%]</td>
</tr>
<tr>
<td>Rate ( \lambda^- )</td>
<td>19.0260(annually)</td>
<td>21.3444(annually)</td>
<td></td>
</tr>
<tr>
<td>Mean Positive</td>
<td>0.2468</td>
<td>0.2513</td>
<td>[-37.67%, 28.83%]</td>
</tr>
<tr>
<td>Jump Size ( \eta^+ )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean Negative</td>
<td>0.5374</td>
<td>0.8213</td>
<td>[-66.04%, 182.14%]</td>
</tr>
<tr>
<td>Jump Size ( \eta^- )</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 3.8: Nasdaq 100: Normalized Model Residues vs Standard Normal

It seems that, in particular during the financial crisis, the SV-DEJ model is better in terms of normality of model residues, especially near zero. The conclusion is reached by comparing standard normal density (solid curve) with the histograms of normalized model residues. Data used is returns of Nasdaq 100 Index, 2001.01-2001.07 for the left three graphs and 2007.08–2011.12 for the right three graphs. From top to bottom, models represented in the graphs are the SV-MJ-JV model, the SV-VG model and the SV-DEJ model.

Left: Before Crisis  Right: During Crisis
Figure 3.9: Nasdaq 100: QQ plots Before and During the Crisis

It appears that the SV-DEJ model provides best fit in both tails among models considered both before and during the financial crisis. In what follow are QQ plots between a sample path simulated from the corresponding model using point estimates in Table 3.9 and actual data (returns of Nasdaq 100 Index, 2001.01-2007.07 for the left three graphs and 2007.08–2011.12 for the right three graphs). From top to bottom, models represented in the graphs are the SV-MJ-JV model, the SV-VG model and the SV-DEJ model. In all of the graphs above, quantiles are compared in the range $-4 \leq Z \leq 4$, where $Z$ is a random variable representing the actual returns.
3.5 Conclusion

We attempt to answer two questions in this chapter: (i) How did jumps in equity returns change during the financial crisis 2007–2011; in particular, were there significant changes in jump rates or in average jump sizes, or both? (ii) In any finite time-horizon, were there finite number of large jumps (e.g. those in affine jump-diffusion models) or infinite number of small jumps (e.g. those in Lévy type models) in equity returns before and during the crisis? To answer these questions, we first found that a simple affine jump-diffusion model with both stochastic volatility and double exponential jump sizes in returns fits both S&P 500 and Nasdaq 100 daily return data well, both before and during the crisis. In fact, the model outperforms existing ones (in particular models with Lévy jumps and affine jump-diffusion models with normal jump sizes) during the crisis, and is at least as good before the crisis. The good fit of the model is in part due to the monotonicity of the exponential distribution, while the normal distribution with negative mean lacks monotonicity. In addition, the heavy-tail feature of the double exponential distribution also helps to fit the data during the crisis.

Based on the model and the data, we answer the aforementioned two questions as follows: (i) We find both positive and negative jump rates increased significantly during the financial crisis, while, on average there seems no significant change of jump sizes; (ii) Our empirical study favors finite number of large jumps in equity returns; in short, affine jump-diffusion model with a proper jump size distribution can fit equity return data well both before and after the crisis.
A.1 Mean-Variance Analysis with Spatial Interaction

Assume that the $n$ returns $\tilde{r} = (r_1, \ldots, r_{n_1}, r_{n_1+1}, \ldots, r_n)'$ satisfy the model (2.2), where the first $n_1$ are ordinary asset returns and the last $n_2$ are futures returns. Then, the mean $\mu$ and covariance matrix $\Sigma$ of $\tilde{r}$ are given by (2.4).

Consider the mean-variance problem faced by an investor in such a market. Let $w$ be the initial wealth of the investor. Let $u = (u_1, \ldots, u_{n_1})'$ denote the vector of dollar-valued wealth invested in the first $n_1$ risky assets and $v = (v_1, \ldots, v_{n_2})'$ denote the vector of dollar-valued positions (i.e., the number of contracts times the futures price) of the investor on the
Define the investor’s portfolio weights as

\[ \phi = (\phi_1, \ldots, \phi_n)' := \frac{1}{w}(u_1, \ldots, u_{n_1}, v_1, \ldots, v_{n_2})'. \]

Then, the net return of the investor’s portfolio is

\[
\begin{align*}
    r_p &= \frac{1}{w} \left( \sum_{i=1}^{n_1} u_i(1 + r_i) + (w - \sum_{i=1}^{n_1} u_i)(1 + r) + \sum_{j=1}^{n_2} v_j (F_{j,1} - F_{j,0}) \right) - 1, \\
    &= \frac{1}{w} \left( \sum_{i=1}^{n_1} u_i(r_i - r) + \sum_{j=1}^{n_2} v_j r_{n_1+j} \right) + r = \phi'(\tilde{r} - r 1_{n_1, n_2}) + r,
\end{align*}
\]

where \(1_{n_1, n_2}\) is defined in (2.8). The mean and variance of \(r_p\) are given by

\[
E[r_p] = h'\phi + r, ~ Var(r_p) = \phi'\Sigma\phi, \text{ where } h = \mu - r 1_{n_1, n_2},
\]

Let \(e\) denote the target mean return of the investor, then the mean-variance problem faced by the investor is

\[
\min_{\phi} \frac{1}{2} \phi'\Sigma\phi \quad \text{s.t.} \quad h'\phi + r = e. \tag{A.1}
\]

Using Lagrange multiplier, we obtain the optimal solution to the problem:

\[
\phi^* = (e - r) \frac{\Sigma^{-1}h}{h'\Sigma^{-1}h}. \tag{A.2}
\]
When there is no risk-free asset in the market, the net return of the investor’s portfolio is

\[ r_p = \frac{1}{w} \left( \sum_{i=1}^{n_1} u_i (1 + r_i) + \sum_{j=1}^{n_2} \frac{v_j}{F_{j,0}} (F_{j,1} - F_{j,0}) \right) - 1 = \phi' \tilde{r}. \]

Then, the mean-variance problem becomes

\[
\min_{\phi} \frac{1}{2} \phi' \Sigma \phi \quad \text{s.t.} \quad \phi' \mu = e \text{ and } \phi' 1_{n_1,n_2} = 1, \tag{A.3}
\]

whose optimal solution can be shown to be

\[ \phi^* = \psi + e \xi, \tag{A.4} \]

where \( \psi = \frac{1}{D} (B \Sigma^{-1} 1_{n_1,n_2} - A \Sigma^{-1} \mu) \), \( \xi = \frac{1}{D} (F \Sigma^{-1} \mu - A \Sigma^{-1} 1_{n_1,n_2}) \), \( A = \mu' \Sigma^{-1} 1_{n_1,n_2} \), \( B = \mu' \Sigma^{-1} \mu \), \( F = 1'_{n_1,n_2} \Sigma^{-1} 1_{n_1,n_2} \), and \( D = BF - A^2 \).

Because both \( \mu \) and \( \Sigma \) are functions of \( \rho \), the optimal portfolio weights \( \phi^* \) and the efficient frontiers are affected by \( \rho \). More specifically, (i) When there exists a risk-free return \( r \), the efficient frontier is \( e = r + \sigma \sqrt{H} \) and \( H \) is a quadratic function of \( \rho \):

\[
H = (\mu - r 1_{n_1,n_2})' \Sigma^{-1} (\mu - r 1_{n_1,n_2}) = (\alpha - r 1_{n_1,n_2})' V^{-1} (\alpha - r 1_{n_1,n_2}) + 2 r 1'_{n_1,n_2} W' V^{-1} (\alpha - r 1_{n_1,n_2}) \rho + r^2 1'_{n_1,n_2} W' V^{-1} W 1_{n_1,n_2} \rho^2. \]

Thus \( H \) is a quadratic function of \( \rho \) and the coefficient in front of \( \rho^2 \) is positive. (ii) When there is no risk-free asset, the effect of spatial interaction on efficient frontier is illustrated in Figure 2.1.
The model parameters used in calculating the efficient frontiers in Figure 2.1 are specified as follows. Let \( n_1 = 10 \) and \( n_2 = 0 \). We first randomly generate 10 points \((x_i, y_i), \ i = 1, 2, \ldots, 10\) on the \( x-y \) plane which denote the locations of 10 assets, assuming that \( \{x_i, y_i : i = 1, \ldots, 10\} \) are i.i.d. normal random variables with mean 0 and variance 100.

We then define the matrix \( W = (w_{ij}) \) as 
\[
    w_{ij} = \frac{1}{s_i d_{ij}} \quad \text{for } i \neq j \quad \text{and} \quad w_{ii} = 0,
\]
where \( d_{ij} \) is the Euclidean distance between asset \( i \) and asset \( j \), and \( s_i := \sum_{k \neq i} d_{ik}^{-1} \). The resulting \( W \) is

\[
    W = \begin{pmatrix}
        0 & 0.080 & 0.131 & 0.206 & 0.054 & 0.055 & 0.128 & 0.068 & 0.204 & 0.075 \\
        0.119 & 0 & 0.082 & 0.193 & 0.067 & 0.055 & 0.223 & 0.069 & 0.103 & 0.089 \\
        0.129 & 0.055 & 0 & 0.086 & 0.073 & 0.093 & 0.066 & 0.122 & 0.273 & 0.103 \\
        0.197 & 0.125 & 0.083 & 0 & 0.048 & 0.044 & 0.261 & 0.056 & 0.119 & 0.068 \\
        0.070 & 0.059 & 0.097 & 0.066 & 0 & 0.119 & 0.056 & 0.212 & 0.093 & 0.228 \\
        0.081 & 0.055 & 0.139 & 0.068 & 0.135 & 0 & 0.058 & 0.235 & 0.107 & 0.122 \\
        0.143 & 0.169 & 0.075 & 0.306 & 0.048 & 0.044 & 0 & 0.054 & 0.096 & 0.064 \\
        0.072 & 0.050 & 0.132 & 0.062 & 0.173 & 0.169 & 0.051 & 0 & 0.108 & 0.181 \\
        0.183 & 0.062 & 0.248 & 0.111 & 0.064 & 0.065 & 0.077 & 0.091 & 0 & 0.099 \\
        0.082 & 0.066 & 0.115 & 0.078 & 0.194 & 0.091 & 0.063 & 0.188 & 0.122 & 0
    \end{pmatrix}.
\]

The vector \( \alpha \) is also a realization of random generation and is given by
\[
    \alpha = (1.334\%, 1.005\%, 1.209\%, 1.141\%, 1.101\%, 1.352\%, 3.531\%, 8.229\%, 1.101\%, 1.893\%)'.
\]

The matrix \( V \) is defined as \( V = 0.015 \cdot I_{10} \) where \( I_{10} \) is a \( 10 \times 10 \) identity matrix.
A.2 Proof for the S-CAPM Theorems

A.2.1 Proof for S-CAPM with Futures

Lemma A.2.1 Let $r_{mv}$ be any mean-variance efficient return (with and without spatial dependence) other than the risk-free return. Then

(i) for any portfolio return $y$ (with and without spatial dependence), it holds that

$$E[y] - r = \frac{Cov(y, r_{mv})}{Var(r_{mv})}(E[r_{mv}] - r);$$

(ii) for any $i = 1, 2, \ldots, n_2$, it holds that

$$E[F_{i,1}] - F_{i,0} = \frac{Cov(F_{i,1}, r_{mv})}{Var(r_{mv})}(E[r_{mv}] - r).$$


Proof. (Proof of Theorem 2.3.1) Suppose that there are $J$ investors in the economy and $w_j$ is the initial wealth of the $j$th investor. Suppose that each investor selects his/her investment portfolio by solving the mean-variance problem (A.1) and the $j$th investor has a target mean return of $e_j$. Then, by (A.2), the position of the $j$th investor is

$$\phi^j = (e_j - r) \frac{\Sigma^{-1} h}{h \Sigma^{-1} h}, j = 1, 2, \ldots, J.$$

Let $C_i$ be the market capitalization of asset $i$, $i = 1, 2, \ldots, n_1$, and $C_M = \sum_{i=1}^{n_1} C_i$ be the
total market capitalization. In market equilibrium, since the aggregate of all positions of futures contracts is zero, it follows that

$$\sum_{j=1}^{J} w_j (e_j - r) h \Sigma^{-1} h = (C_1, \ldots, C_{n_1}, 0, \ldots, 0)'.$$

(A.5)

Hence,

$$\frac{\Sigma^{-1} h}{h' \Sigma^{-1} h} = \begin{pmatrix} g \\ 0 \end{pmatrix}, \text{ where } g = \frac{1}{\sum_{j=1}^{J} w_j (e_j - r)} (C_1, \ldots, C_{n_1})'.$$

Therefore, in equilibrium, each investor holds only the market portfolio and the risk-free asset, and no investor trades the futures contracts.

Furthermore, let $e_M = \sum_{j=1}^{J} w_j e_j / C_M$. Then, it follows from (A.5) that

$$(C_1, \ldots, C_{n_1}, 0, \ldots, 0)' = C_M (e_M - r) \frac{\Sigma^{-1} h}{h' \Sigma^{-1} h},$$

which shows that the market portfolio is mean-variance efficient. Then, the conclusion in Theorem 2.3.1 follows by applying Lemma A.2.1 with the market portfolio return $r_M$ being $r_{mv}$.

A.2.2 S-CAPM with Futures When There Is No Risk-free Asset

Theorem A.2.1 (S-CAPM with Futures When There Is No Risk-free Asset) Suppose that there is no risk free asset and that the returns of $n = n_1 + n_2$ risky assets are generated by the model (2.2), of which the first $n_1$ are returns of ordinary assets and the others are
defined in (2.5) which are “nominal returns” of futures contracts. Suppose that \( n_1 > 0 \).

Let \( r_M \) be the return of market portfolio. If each investor holds mean-variance efficient portfolio, then in equilibrium, \( r_M \) is mean-variance efficient. Furthermore, if \( r_M \) is not the minimum-variance return, then there exists another mean-variance efficient return \( r_0 \) such that \( \text{Cov}(r_M, r_0) = 0 \), and it holds that

(i) for the ordinary assets,

\[
E[r_i] - E[r_0] = \frac{\text{Cov}(r_i, r_M)}{\text{Var}(r_M)}(E[r_M] - E[r_0]), \quad i = 1, 2, \ldots, n_1;
\]

(ii) for the futures contracts,

\[
E[F_{i,1}] - F_{i,0} = \frac{\text{Cov}(F_{i,1}, r_M)}{\text{Var}(r_M)}(E[r_M] - E[r_0]), \quad i = 1, 2, \ldots, n_2.
\]

**Lemma A.2.2** Let \( r_{mv} \) be any mean-variance efficient return other than the minimum-variance return. Then, there exists another mean-variance efficient return \( r_0 \) such that \( \text{Cov}(r_{mv}, r_0) = 0 \). Furthermore, it holds that

(i) for any portfolio return \( y \),

\[
E[y] - E[r_0] = \frac{\text{Cov}(y, r_{mv})}{\text{Var}(r_{mv})}(E[r_{mv}] - E[r_0]);
\]
(ii) for futures contracts,

\[ E[F_{i,1}] - F_{i,0} = \frac{\text{Cov}(F_{i,1}, r_{mv})}{\text{Var}(r_{mv})}(E[r_{mv}] - E[r_0]), \quad i = 1, 2, \ldots, n_2. \]

**Proof.** See Kou, Peng, and Zhong (2013). \(\square\)

**Proof.** (Proof of Theorem A.2.1.) When there is no risk-free asset, the mean-variance problem faced by an investor is given by (A.3), and the optimal portfolio weight is given by (A.4). Suppose that there are \(J\) investors in the economy and let \(w_j\) and \(e_j\) be the initial wealth and target mean return of the \(j\)th investor. Then, the position of the \(j\)th investor is

\[ \phi^j = \psi + e_j \xi, \quad j = 1, 2, \ldots, J. \]

Let \(C_i\) be the market capitalization of asset \(i, i = 1, 2, \ldots, n_1, \) and \(C_M = \sum_{i=1}^{n_1} C_i\) be the total market capitalization. In market equilibrium, since the aggregate of all positions of futures contracts is zero, it follows that

\[ \sum_{j=1}^{J} w_j \phi^j = \left( \sum_{j=1}^{J} w_j \right) \psi + \left( \sum_{j=1}^{J} w_j e_j \right) \xi = (C_1, \ldots, C_{n_1}, 0, \ldots, 0)'. \tag{A.6} \]

Let \(e_M = \sum_{j=1}^{J} w_j e_j / C_M.\) Then, since \(C_M = \sum_{j=1}^{J} w_j,\) it follows from (A.6) that

\[ \psi + e_M \xi = \frac{1}{C_M} (C_1, \ldots, C_{n_1}, 0, \ldots, 0)', \]

81
which shows that the market portfolio is mean-variance efficient with a target mean return $e_M$. Then, the conclusion in Theorem A.2.1 follows by applying Lemma A.2.2 with the market portfolio return $r_M$ being $r_{mv}$.

\[ \square \]

A.3 Proof for the S-APT Theorems

A.3.1 Proof for Theorem 2.4.1

**Proof.** Fix any $\delta > 0$. Without loss of generality, assume that $|\bar{\alpha}_j^{(n)}| > \delta$, $j = 1, \ldots, N(n, \delta)$.

We rewrite (2.13) as

\[ (I_n - \rho^{(n)} W^{(n)})(\bar{r}^{(n)} - r_{1n_1,n_2}) = \bar{\alpha}^{(n)} + B^{(n)} \tilde{f} + \tilde{\epsilon}^{(n)}. \]

Let $\eta_j$ be the $j$th column of $I_n$. For $1 \leq j \leq N(n, \delta)$, if $\bar{\alpha}_j^{(n)} > \delta$, consider the zero-cost portfolio with payoff $\eta_j'(I_n - \rho^{(n)} W^{(n)})(\bar{r}^{(n)} - r_{1n_1,n_2}) - \eta_j' B^{(n)} \tilde{F}$, which by definition is equal to $\bar{\alpha}_j^{(n)} + \epsilon_j^{(n)}$, a random variable with mean $\bar{\alpha}_j^{(n)} > \delta$ and variance not exceeding $\sigma^2$; if $\bar{\alpha}_j^{(n)} < -\delta$, one can construct another zero-cost portfolio with payoff $-\bar{\alpha}_j^{(n)} - \epsilon_j^{(n)}$ by taking opposite positions of the previous portfolio. In this way, $N(n, \delta)$ such portfolios can be constructed. Since the components of $\tilde{\epsilon}^{(n)}$ are uncorrelated with each other, a portfolio with equal weights in these $N(n, \delta)$ portfolios has a payoff with mean greater than $\delta$ and variance less than $\sigma^2 / N(n, \delta)$. If there exists a subsequence $\{m_1, m_2, \ldots\}$ such that $N(m_k, \delta)$ grows unboundedly as $k$ goes to infinity, then the corresponding sequence of portfolios will have
payoffs with means greater than $\delta$ and diminishing variances, constituting an asymptotic arbitrage opportunity. Therefore, if no asymptotic arbitrage opportunities exist, then there exists a number $N_\delta$ not depending on $n$ such that $N(n, \delta) < N_\delta$ for all $n$. Since $\delta$ can be arbitrarily small, we conclude that $\bar{\alpha}^{(n)} \approx 0$. □

A.3.2 Proof for Theorem 2.4.2

**Proof.** Since $I_n - \rho^{(n)} W^{(n)}$ is invertible, the model (2.15) can be written as

$$\bar{r}^{(n)} = Q^{(n)} \alpha^{(n)} + Q^{(n)} B^{(n)} \bar{f} + Q^{(n)} \bar{\epsilon}^{(n)}; \text{ where } Q^{(n)} = (I_n - \rho^{(n)} W^{(n)})^{-1}. \quad (A.7)$$

For the sake of notational simplicity, the superscript $(n)$ will be dropped when the meaning is clear. Let

$$\hat{\alpha} = Q \alpha, \hat{B} = QB, \hat{\epsilon} = Q \bar{\epsilon}, \Omega = QQ'. \quad (A.8)$$

Since $\Omega$ is positive definite, it can be factored as $\Omega = CC'$ where $C$ is a nonsingular matrix. Project $C^{-1} \hat{\alpha}$ into the space spanned by $C^{-1} 1_{n_1, n_2}$ and the columns of $C^{-1} \hat{B}$:

$$C^{-1} \hat{\alpha} = C^{-1} 1_{n_1, n_2} \lambda_0 + C^{-1} \hat{B} \lambda + u.$$
Define the pricing errors \( v := \hat{\alpha} - 1_{n_1,n_2} \lambda_0 - \hat{B} \lambda = Cu. \) Then, by orthogonality, we have

\[
0 = \hat{B}'(C')^{-1}u = \hat{B}'(C')^{-1}C^{-1}v = \hat{B}'\Omega^{-1}v, \tag{A.9}
\]

\[
0 = 1'_{n_1,n_2}(C')^{-1}u = 1'_{n_1,n_2}(C')^{-1}C^{-1}v = 1'_{n_1,n_2}\Omega^{-1}v. \tag{A.10}
\]

Consider the portfolio \( h = \Omega^{-1}v(v'\Omega^{-1}v)^{-1}. \) By (A.10), \( h \) is a zero-cost portfolio. By (A.9), the payoff of the zero-cost portfolio is

\[
h'\tilde{r} = h'(Q\alpha + QB\tilde{f} + Q\tilde{\epsilon}) = h'(\hat{\alpha} + \hat{B}\tilde{f} + \tilde{\epsilon}) = h'\hat{\alpha} + (v'\Omega^{-1}v)^{-1}v'\Omega^{-1}\hat{B}\tilde{f} + h'\tilde{\epsilon}
\]

\[
= h'\hat{\alpha} + h'\tilde{\epsilon},
\]

whose expectation and variance are

\[
E[h'\tilde{r}] = h'\hat{\alpha} = (v'\Omega^{-1}v)^{-1}v'\Omega^{-1}(1_{n_1,n_2}\lambda_0 + \hat{B}\lambda + v) = 1,
\]

\[
Var(h'\tilde{r}) = h'\Omega h = (v'\Omega^{-1}v)^{-2}v'\Omega^{-1}\Omega\Omega^{-1}v = (v'\Omega^{-1}v)^{-1}
\]

\[
= [(\hat{\alpha} - 1_{n_1,n_2}\lambda_0 - \hat{B}\lambda)'(Q')^{-1}V^{-1}(Q)^{-1}(\hat{\alpha} - 1_{n_1,n_2}\lambda_0 - \hat{B}\lambda)]^{-1}
\]

\[
= [(\alpha - (I_n - \rho W)1_{n_1,n_2}\lambda_0 - B\lambda)'V^{-1}(\alpha - (I_n - \rho W)1_{n_1,n_2}\lambda_0 - B\lambda)]^{-1}
\]

\[
= (U'V^{-1}U)^{-1}.
\]

Therefore, if (2.24) is violated, the variance of \( h'\tilde{r} \) vanishes along some subsequence, which constitutes an asymptotic arbitrage opportunity.
The proof for the case when there exists a risk-free return $r$ is almost a copy of the above. Let $Q, \hat{\alpha}, \hat{B}, \tilde{\epsilon},$ and $\Omega$ be defined in (A.7) and (A.8). Since $\Omega$ is positive definite, it can be factored as $\Omega = CC'$ where $C$ is a nonsingular matrix. Project $C^{-1}(\hat{\alpha} - r1_{n_1,n_2})$ onto the space spanned by the columns of $C^{-1}\hat{B}$:

$$C^{-1}(\hat{\alpha} - r1_{n_1,n_2}) = C^{-1}\hat{B}\lambda + u.$$

Define the pricing errors $v := \hat{\alpha} - r1_{n_1,n_2} - \hat{B}\lambda = Cu$. Then, by orthogonality, we have

$$0 = \hat{B}'(C')^{-1}u = \hat{B}'(C')^{-1}C^{-1}v = \hat{B}'\Omega^{-1}v. \quad \text{(A.11)}$$

Consider the zero-cost portfolio which has dollar-valued positions $h = (v'\Omega^{-1}v)^{-1}\Omega^{-1}v$ in the $n_1$ risky assets and the $n_2$ futures contracts and the position $-h'1_{n_1,n_2}$ in the risk-free asset. By (A.8) and (A.11), the payoff of the portfolio is

$$h'(1_{n_1,n_2} + \tilde{r}) - h'1_{n_1,n_2}(1 + r) = h'(\tilde{r} - r1_{n_1,n_2}) = h'(Q\alpha + QB\tilde{f} + Q\tilde{\epsilon} - r1_{n_1,n_2})$$

$$= h'(\hat{\alpha} + \hat{B}\tilde{f} + \tilde{\epsilon} - r1_{n_1,n_2}) = h'(\hat{\alpha} - r1_{n_1,n_2}) + (v'\Omega^{-1}v)^{-1}v'\Omega^{-1}\hat{B}\tilde{f} + h'\tilde{\epsilon}$$

$$= h'(\hat{\alpha} - r1_{n_1,n_2}) + h'\tilde{\epsilon},$$

85
whose mean and variance are

\[
E[h'(\tilde{r} - r1_{n1,n2})] = h'(\hat{\alpha} - r1_{n1,n2}) = (v'\Omega^{-1}v)^{-1}v'\Omega^{-1}(\hat{B}\lambda + v) = 1,
\]

\[
\text{Var}(h'(\tilde{r} - r1_{n1,n2})) = \text{Var}(h'\tilde{\varepsilon}) = h'\Omega h = (v'\Omega^{-1}v)^{-2}v'\Omega^{-1}\Omega\Omega^{-1}v
\]

\[
= (v'\Omega^{-1}v)^{-1} = [(\hat{\alpha} - r1_{n1,n2} - \hat{B}\lambda)'(Q')^{-1}V^{-1}(Q)^{-1}(\hat{\alpha} - r1_{n1,n2} - \hat{B}\lambda)]^{-1}
\]

\[
= [(\alpha - (I_n - \rho W)1_{n1,n2}r - B\lambda)'V^{-1}(\alpha - (I_n - \rho W)1_{n1,n2}r - B\lambda)]^{-1}.
\]

Therefore, if (2.24) (with \(\lambda_0^{(n)}\) replaced by \(r\)) is violated, then the variance of \(h'(\tilde{r} - r1_{n1,n2})\) vanishes along some subsequence, and an asymptotic arbitrage opportunity exists. The proof is thus completed. \(\square\)

### A.4 Proof for Proposition 2.5.1

**Proof.** \(\theta_0\) is not identifiable if and only if there exists \(\theta \in \Theta\) and \(\theta \neq \theta_0\) such that

\[
P(l(\tilde{y}_t \mid \tilde{g}_t; \theta) = l(\tilde{y}_t \mid \tilde{g}_t; \theta_0)) = 1. \tag{A.12}
\]

It follows from the part (ii) of Assumption 2.5.1 that \(P((\tilde{y}_t', \tilde{g}_t')' \in \mathbb{R}^n \times \mathbb{A}) > 0\) and the joint distribution of \((\tilde{y}_t', \tilde{g}_t')'\) has a strictly positive density on \(\mathbb{R}^n \times \mathbb{A}\). Therefore, (A.12) implies that

\[
l(\tilde{y} \mid \tilde{g}; \theta) = l(\tilde{y} \mid \tilde{g}; \theta_0), \quad \forall(\tilde{y}, \tilde{g}) \in \mathbb{R}^n \times \mathbb{A}.
\]
By (2.34), we have

$$l(\tilde{y} \mid \tilde{g}; \theta) = \tilde{y}' C_1(\theta) \tilde{y} + \tilde{g}' C_2(\theta) \tilde{g} + \tilde{y}' C_3(\theta) \tilde{y} + C_4(\theta)' \tilde{y} + C_5(\theta)' \tilde{g} + C_6(\theta), \quad (A.13)$$

where

$$C_1(\theta) = -\frac{1}{2\sigma^2} (I_n - \rho W')(I_n - \rho W)$$
$$C_2(\theta) = -\frac{1}{2\sigma^2} B' B \quad (A.14)$$
$$C_3(\theta) = \frac{1}{\sigma^2} B' (I_n - \rho W) \quad (A.15)$$
$$C_4(\theta) = \frac{1}{\sigma^2} (I_n - \rho W') \bar{\alpha} \quad (A.16)$$
$$C_5(\theta) = -\frac{1}{\sigma^2} B' \bar{\alpha} \quad (A.17)$$
$$C_6(\theta) = -\frac{n}{2} \log(2\pi \sigma^2) + \frac{1}{2} \log(\det((I_n - \rho W')(I_n - \rho W))) - \frac{1}{2\sigma^2} \bar{\alpha}' \bar{\alpha}. \quad (A.18)$$

By the equality of partial derivatives of $l(\tilde{y} \mid \tilde{g}, \theta)$ and $l(\tilde{y} \mid \tilde{g}, \theta_0)$ on $\mathbb{R}^n \times A$, we obtain that

$$C_i(\theta) = C_i(\theta_0), \quad i = 1, 2, \ldots, 6. \quad (A.19)$$

Since $I_n - \rho W$ is invertible, (A.15) and (A.16) imply that

$$B' = \sigma^2 C_3(\theta) (I_n - \rho W)^{-1}, \quad \bar{\alpha} = \sigma^2 (I_n - \rho W')^{-1} C_4(\theta). \quad (A.20)$$
Then, plugging (A.20) into (A.14), (A.17), and (A.18), we obtain

\[ C_2(\theta) = \frac{1}{4} C_3(\theta) C_1(\theta)^{-1} C_3(\theta)', \]
\[ C_5(\theta) = \frac{1}{2} C_3(\theta) C_1(\theta)^{-1} C_4(\theta), \]
\[ C_6(\theta) = -\log((2\pi)^{\frac{3}{2}} \det(-2C_1(\theta))^{-\frac{1}{2}}) + \frac{1}{4} C_4(\theta)'C_1(\theta)^{-1}C_4(\theta). \]

Hence, (A.19) is equivalent to

\[ C_i(\theta) = C_i(\theta_0), \; i = 1, 3, 4. \]  (A.21)

Now we are ready to prove the proposition. We will first show the sufficiency.

(i) Suppose that \( W \) is regular. Suppose for the sake of contradiction that there exists \( \theta \neq \theta_0 \) such that (A.12) holds. Then, (A.21) holds. \( C_1(\theta) = C_1(\theta_0) \) is equivalent to that

\[ \left( \frac{1}{\sigma_0^2} - \frac{1}{\sigma^2} \right) I_n - \left( \frac{\rho_0}{\sigma_0^2} - \frac{\rho}{\sigma^2} \right) (W' + W) + \left( \frac{\rho_0^2}{\sigma_0^2} - \frac{\rho^2}{\sigma^2} \right) W'W = 0. \]  (A.22)

Considering the \((i, i)\)-element of the matrix on the left and noting \( W_{ii} = 0 \), we obtain

\[ \frac{1}{\sigma_0^2} - \frac{1}{\sigma^2} + \left( \frac{\rho_0^2}{\sigma_0^2} - \frac{\rho^2}{\sigma^2} \right) \sum_{k=1}^{n} W_{ki}^2 = 0, \; i = 1, 2, \ldots, n. \]  (A.23)
For $i < j$, considering the $(i, j)$-element of the matrices on both sides of (A.22), we obtain

$$-(\frac{\rho_0}{\sigma_0^2} - \frac{\rho}{\sigma^2})(W_{ij} + W_{ji}) + \left(\frac{\rho_0^2}{\sigma_0^2} - \frac{\rho^2}{\sigma^2}\right) \sum_{k=1}^{n} W_{ki}W_{kj} = 0, \forall 1 \leq i < j \leq n. \quad (A.24)$$

Suppose that $\frac{\rho_0^2}{\sigma_0^2} - \frac{\rho^2}{\sigma^2} = 0$, then (A.23) implies $\sigma = \sigma_0$, which, together with (A.24) and that there exists $i \neq j$ such that $W_{ij} + W_{ji} > 0$ (by Assumption 2.5.2), implies that $\rho = \rho_0$. Then, since $I_n - \rho W$ is invertible, $C_5(\theta) = C_3(\theta_0)$ and (A.15) imply $B = B_0$; $C_4(\theta) = C_4(\theta_0)$ and (A.16) imply that $\bar{\alpha} = \bar{\alpha}_0$. Therefore, we have shown that $\theta = \theta_0$, but this contradicts to the assumption that $\theta \neq \theta_0$. Hence, $\frac{\rho_0^2}{\sigma_0^2} \neq \frac{\rho^2}{\sigma^2}$. Suppose without generality that $\frac{\rho_0^2}{\sigma_0^2} > \frac{\rho^2}{\sigma^2}$. Since (A.23) and that there exists $i$ such that $\sum_{k=1}^{n} W_{ki}^2 > 0$ (by Assumption 2.5.2), it follows that $\sigma_0 > \sigma$. Hence, (A.23) and (A.24) imply that

$$\sum_{k=1}^{n} W_{ki}^2 = c_1, \quad i = 1, 2, \ldots, n; \text{ and } \sum_{k=1}^{n} W_{ki}W_{kj} = c_2(W_{ij} + W_{ji}), \forall 1 \leq i < j \leq n;$$

where

$$c_1 = -\frac{1}{\sigma_0^2} - \frac{1}{\sigma^2} > 0 \quad \text{and} \quad c_2 = \frac{\rho_0^2}{\sigma_0^2} - \frac{\rho^2}{\sigma^2} \geq 0, \quad (A.25)$$

which contradicts to the assumption that $W$ is regular.

(ii) Suppose that $W$ is not regular and corresponds to $c_1 > 0$ and $c_2 \geq 0$ in (2.35) and (2.36). Suppose for the sake of contradiction that there exists $\theta \neq \theta_0$ such that (A.12) holds. Then, by the same argument in case (i) and by the uniqueness of $(c_1, c_2)$, $\sigma \neq \sigma_0$ and $(\rho, \sigma)$ must satisfy the equations in (A.25) and hence must be a solution to the following
two equations

\[ c_1 \left( \frac{\rho_0^2}{\sigma_0^2} - \frac{\rho^2}{\sigma^2} \right) = -\frac{1}{\sigma_0^2} + \frac{1}{\sigma^2} \quad \text{and} \quad c_2 \left( \frac{\rho_0^2}{\sigma_0^2} - \frac{\rho^2}{\sigma^2} \right) = \frac{\rho_0}{\sigma_0^2} - \frac{\rho}{\sigma^2}. \]

It can be shown by simple algebra that the above two equations are equivalent to

\[ (c_2 + c_1 \rho_0) \rho^2 - (c_1 \rho_0^2 + 1) \rho - (c_2 \rho_0^2 - \rho_0) = 0 \quad \text{and} \quad \sigma^2 = \frac{1 + c_1 \rho_0^2}{1 + c_1 \rho_0} \sigma_0^2. \quad (A.26) \]

If \( c_2 + c_1 \rho_0 = 0 \), then the above system of equations has a unique solution \((\rho_0, \sigma_0)\); otherwise, the above equations have two solutions \((\rho_0, \sigma_0)\) and \( \left( \frac{1-c_2 \rho_0}{c_2 + c_1 \rho_0}, \sigma_0 \sqrt{\frac{c_2^2 + c_1}{(c_2 + c_1 \rho_0)^2}} \right) \).

(ii.1) Suppose that (2.37) or (2.38) holds, then the two equations in (A.26) have a unique solution \((\rho_0, \sigma_0)\), and hence the two equations in (A.25) do not have a solution \((\rho, \sigma) \neq (\rho_0, \sigma_0)\), which leads to a contradiction.

(ii.2) Suppose that (2.39) holds, then \( \left( \frac{1-c_2 \rho_0}{c_2 + c_1 \rho_0}, \sigma_0 \sqrt{\frac{c_2^2 + c_1}{(c_2 + c_1 \rho_0)^2}} \right) \) is the unique solution to (A.25); hence, \( \rho = \frac{1-c_2 \rho_0}{c_2 + c_1 \rho_0} = \rho_\ast \) and \( \sigma = \sigma_0 \sqrt{\frac{c_2^2 + c_1}{(c_2 + c_1 \rho_0)^2}} = \sigma_\ast \). Since \( C_3(\theta) = C_3(\theta_0) \) and \( C_4(\theta) = C_4(\theta_0) \), it follows that \( \bar{\alpha} = \bar{\alpha}_\ast \) and \( B = B_\ast \). Hence, \( \theta = \theta_\ast \). However, \( \theta \in \Theta \) but \( \theta_\ast \notin \Theta \), which constitutes a contradiction.

Therefore, we have completed the proof of sufficiency. We will then show the necessity. Suppose for the sake of contradiction that \( W \) does not satisfy any of the conditions specified. Then, \( W \) is not regular, and it corresponds to a unique pair of \( c_1 > 0 \) and \( c_2 \geq 0 \) and \( \rho_0 \neq -\frac{c_2}{c_1} \), and \( \frac{1-c_2 \rho_0}{c_2 + c_1 \rho_0} \neq \rho_0 \), and \( \theta_\ast \in \Theta \). Then, by the definition of \( \theta_\ast \), it holds that \( \theta_\ast \neq \theta_0 \) and \( C_i(\theta_\ast) = C_i(\theta_0) \) for \( i = 1, 3, \) and \( 4 \), which further implies that \( C_i(\theta_\ast) = C_i(\theta_0) \)
for $i = 1, 2, \ldots, 6$. Therefore, $l(\tilde{y}_t \mid \tilde{g}_t, \theta^*) = l(\tilde{y}_t \mid \tilde{g}_t, \theta_0)$, but this contradicts to that $\theta_0$ is identifiable.

\section*{A.5 Proof for Proposition 2.5.2}

\textbf{Proof.} Since $\det((I_n - \rho W')(I_n - \rho W))$ is equal to a polynomial of $\rho$, it follows from (A.13) that $\hat{Q}_T(\theta)$ is twice continuously differentiable on the interior of $\Theta$. The proof for part (i) to (vi) is as follows.

(i) Let $f(\tilde{y}_t \mid \tilde{g}_t, \theta)$ denote the conditional density. It follows from the model (2.29) and part (i) of Assumption 2.5.1 that $E[\|\tilde{y}_t\|^2] < \infty$. Hence, (A.13) implies that for any $\theta \in \Theta$, $E[l(\tilde{y}_t \mid \tilde{g}_t, \theta)] < \infty$. For any $\theta \neq \theta_0$, define $g(\tilde{y}_t, \tilde{g}_t) := \frac{f(\tilde{y}_t \mid \tilde{g}_t, \theta)}{f(\tilde{y}_t \mid \tilde{g}_t, \theta_0)}$. Since $\theta_0$ is identifiable, it follows that $P(g(\tilde{y}_t, \tilde{g}_t) \neq 1) > 0$. Therefore, it follows from the strict Jensen’s inequality that

$$E[l(\tilde{y}_t \mid \tilde{g}_t, \theta_0) - l(\tilde{y}_t \mid \tilde{g}_t, \theta)] = E[-\log g(\tilde{y}_t, \tilde{g}_t)] > -\log E[g(\tilde{y}_t, \tilde{g}_t)]. \tag{A.27}$$

Since

$$E[g(\tilde{y}_t, \tilde{g}_t) \mid \tilde{g}_t] = \int \frac{f(\tilde{y}_t \mid \tilde{g}_t, \theta)}{f(\tilde{y}_t \mid \tilde{g}_t, \theta_0)} f(\tilde{y}_t \mid \tilde{g}_t, \theta_0) d\tilde{y}_t = \int f(\tilde{y}_t \mid \tilde{g}_t, \theta) d\tilde{y}_t = 1,$$

it follows that $E[g(\tilde{y}_t, \tilde{g}_t)] = 1$, which in combination with (A.27) implies that $Q_0(\theta)$ has a unique maximizer $\theta_0$. 

\vspace{10pt}

91
(ii) and (iii). We first show that

$$E[\sup_{\theta \in \Theta} |l(\tilde{y}_t \mid \tilde{g}_t; \theta)|] < \infty.$$  
(A.28)

By (A.13),

$$l(\tilde{y}_t \mid \tilde{g}_t; \theta) = \sum_{i,j} a_{ij}(\theta)y_{it}y_{jt}$$

$$+ \sum_{i,j} b_{ij}(\theta)g_{it}g_{jt} + \sum_{i,j} c_{ij}(\theta)y_{it}g_{jt} + \sum_{i=1}^{n} d_i(\theta)y_{it} + \sum_{i=1}^{K} e_i(\theta)g_{it} + C_6(\theta),$$

where $a_{ij}(\cdot)$, $b_{ij}(\cdot)$, $c_{ij}(\cdot)$, $d_i(\cdot)$, $e_i(\cdot)$, and $C_6(\cdot)$ are all continuous functions. Since $\Theta$ is compact, it follows that

$$E[\sup_{\theta \in \Theta} |a_{ij}(\theta)y_{it}y_{jt}|] = E[|y_{it}y_{jt}|] \sup_{\theta \in \Theta} |a_{ij}(\theta)| < \infty.$$ 

Similarly, the expectation of the supremum (with respect to $\theta$) of the absolute value of each term in the summation for $l(\tilde{y}_t \mid \tilde{g}_t; \theta)$ is finite; therefore, $E[\sup_{\theta \in \Theta} |l(\tilde{y}_t \mid \tilde{g}_t; \theta)|] < \infty$. Then, since $l(\tilde{y}_t \mid \tilde{g}_t, \theta)$ is continuous at every $\theta \in \Theta$, it follows from Lemma 2.4 in Neway and McFadden (1994, p. 2129) that (ii) and (iii) hold.

(iv) Define $\tilde{\xi}_t := \tilde{y}_t - \rho W\tilde{y}_t - X_t \hat{\beta}$. By Jacobi’s formula of matrix calculus,

$$\frac{d}{d\rho} \det((I_n - \rho W')(I_n - \rho W)) = -2 \det((I_n - \rho W')(I_n - \rho W)) \text{tr}((I_n - \rho W)^{-1}W),$$
where $\text{tr}(\cdot)$ denotes the trace of a matrix. Hence,

$$
\frac{\partial l(\tilde{y}_t | \tilde{g}_t, \theta)}{\partial \rho} = -\text{tr}((I_n - \rho W)^{-1}W) + \frac{1}{\sigma^2} \tilde{\xi}_t' W \tilde{y}_t. \tag{A.29}
$$

By simple algebra, we have

$$
\frac{\partial l(\tilde{y}_t | \tilde{g}_t, \theta)}{\partial b} = \frac{1}{\sigma^2} X_t' \tilde{\xi}_t, \quad \frac{\partial l(\tilde{y}_t | \tilde{g}_t, \theta)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \tilde{\xi}_t' \tilde{\xi}_t. \tag{A.30}
$$

Hence,

$$
E\left[ \frac{\partial l(\tilde{y}_t | \tilde{g}_t, \theta_0)}{\partial \rho} \right] = -\text{tr}((I_n - \rho_0 W)^{-1}W) + \frac{1}{\sigma_0^2} E[\tilde{\xi}_t' W(I_n - \rho_0 W)^{-1}(\tilde{\alpha}_0 + B_0 \tilde{g}_t + \tilde{\xi}_t)]
$$

$$
= -\text{tr}((I_n - \rho_0 W)^{-1}W) + \frac{1}{\sigma_0^2} E[\tilde{\xi}_t' W(I_n - \rho_0 W)^{-1}\tilde{\xi}_t]
$$

$$
= -\text{tr}((I_n - \rho_0 W)^{-1}W) + \frac{1}{\sigma_0^2} E[\text{tr}(W(I_n - \rho_0 W)^{-1}\tilde{\xi}_t' \tilde{\xi}_t)]
$$

$$
= -\text{tr}((I_n - \rho_0 W)^{-1}W) + \frac{1}{\sigma_0^2} \sigma_0^2 \text{tr}(W(I_n - \rho_0 W)^{-1}) = 0.
$$

Also, by (A.30),

$$
E\left[ \frac{\partial l(\tilde{y}_t | \tilde{g}_t, \theta_0)}{\partial b} \right] = E\left[ \frac{1}{\sigma_0^2} X_t' \tilde{\xi}_t \right] = 0, \quad E\left[ \frac{\partial l(\tilde{y}_t | \tilde{g}_t, \theta_0)}{\partial \sigma^2} \right] = E\left[ -\frac{n}{2\sigma_0^2} + \frac{1}{2\sigma_0^4} \tilde{\xi}_t' \tilde{\xi}_t \right] = 0,
$$

which completes the proof for part (iv).

(v) For brevity of notation, we use $l(\theta)$ to denote $l(\tilde{y}_t | \tilde{g}_t, \theta)$ in the sequel. By (A.29)
and (A.30), we have

\[
\frac{\partial^2 l(\theta)}{\partial \rho^2} = -\text{tr}(W(I_n - \rho W)^{-1}W(I_n - \rho W)^{-1}) - \frac{1}{\sigma^2} \tilde{y}_t W' W \tilde{y}_t, \quad (A.31)
\]

\[
\frac{\partial^2 l(\theta)}{\partial \rho \partial b} = -\frac{1}{\sigma^2} X_t' W \tilde{y}_t, \quad \frac{\partial^2 l(\theta)}{\partial \sigma^2 \partial b} = -\frac{1}{\sigma^2} X_t' X_t, \quad \frac{\partial^2 l(\theta)}{\partial \rho \partial \sigma^2} = -\frac{1}{\sigma^4} \tilde{\xi}_t^2 W \tilde{y}_t, \quad (A.32)
\]

\[
\frac{\partial^2 l(\theta)}{\partial \sigma^2 \partial \sigma^2} = -\frac{1}{\sigma^4} X_t' \tilde{\xi}_t, \quad (A.33)
\]

which completes the proof.

(vi) Let

\[ C := W (I_n - \rho_0 W)^{-1} \quad (A.34) \]

and let \( C_{ij} \) be the \((i, j)\) element of \( C \). Then by (A.31), we have

\[
E\left[ \frac{\partial^2 l(\theta_0)}{\partial \rho^2} \mid \tilde{g}_t \right] = -\text{tr}(C^2) - \frac{1}{\sigma^2} E[((I_n - \rho_0 W)^{-1}(X_t\bar{b}_0 + \tilde{\epsilon}_t))'W'(I_n - \rho_0 W)^{-1}(X_t\bar{b}_0 + \tilde{\epsilon}_t) \mid \tilde{g}_t]
\]

\[
= -\text{tr}(C^2) - \frac{1}{\sigma^2} b_0' X_t' C' C X_t \bar{b}_0 - \frac{1}{\sigma^2} E[\tilde{\epsilon}_t' C' C \tilde{\epsilon}_t]
\]

\[
= -\text{tr}(C^2) - \frac{1}{\sigma^2} b_0' X_t' C' C X_t \bar{b}_0 - \text{tr}(C' C). \quad (A.35)
\]

By (A.29) and simple algebra,

\[
E\left[ \left( \frac{\partial l(\theta_0)}{\partial \rho} \right)^2 \mid \tilde{g}_t \right] = E[(- \text{tr}(C) + \frac{1}{\sigma^2} \tilde{\epsilon}_t' W \tilde{y}_t)^2 \mid \tilde{g}_t]
\]

\[
= E[(- \text{tr}(C) + \frac{1}{\sigma^2} \tilde{\epsilon}_t' C(X_t\bar{b}_0 + \tilde{\epsilon}_t))^2 \mid \tilde{g}_t]
\]

94
\[ E[\text{tr}(C)^2 - \frac{2\text{tr}(C)}{\sigma_0^2} \tilde{\epsilon}_t^i C(X_t b_0 + \tilde{\epsilon}_t) | \tilde{g}_t] + E[\frac{1}{\sigma_0^2} \tilde{\epsilon}_t^i C(X_t b_0 + \tilde{\epsilon}_t)(X_t b_0 + \tilde{\epsilon}_t)'C' \tilde{\epsilon}_t | \tilde{g}_t] = -\text{tr}(C)^2 + \frac{1}{\sigma_0^2} \tilde{b}_0' X_t' C X_t b_0 - \frac{2}{\sigma_0^4} \tilde{b}_0' X_t' C' \epsilon X_t C \tilde{\epsilon}_t + \frac{1}{\sigma_0^4} E[(\epsilon_t C \tilde{\epsilon}_t)^2] \]

\[ = -\text{tr}(C)^2 + \frac{1}{\sigma_0^2} \tilde{b}_0' X_t' C' C X_t b_0 + \frac{1}{\sigma_0^4} E[(\epsilon_t C \tilde{\epsilon}_t)^2] \]

\[ = -\text{tr}(C)^2 + \frac{1}{\sigma_0^2} \tilde{b}_0' X_t' C' C X_t b_0 + \frac{1}{\sigma_0^2} E[(\epsilon_t C \tilde{\epsilon}_t)^2] \]

\[ = -\text{tr}(C)^2 + \frac{1}{\sigma_0^2} \tilde{b}_0' X_t' C' C X_t b_0 + \frac{1}{\sigma_0^4} E[(\sum_{i=1}^n \sum_{j=1}^n C_{ij} \epsilon_{it} \epsilon_{jt})^2] \]

\[ = - (\sum_{i=1}^n C_{ii})^2 + \frac{1}{\sigma_0^2} \tilde{b}_0' X_t' C' C X_t b_0 + \frac{1}{\sigma_0^4} E[(\sum_{i=1}^n C_{ii} \epsilon_{it}^2 + \sum_{i<j} (C_{ij} + C_{ji}) \epsilon_{it} \epsilon_{jt})^2] \]

\[ = 2 \sum_{i=1}^n C_{ii}^2 + \sum_{i \neq j} C_{ij}^2 + 2 \sum_{i<j} C_{ij} C_{ji} + \frac{1}{\sigma_0^2} \tilde{b}_0' X_t' C' C X_t b_0 \]

\[ = -E[\frac{\partial^2 l(\theta_0)}{\partial \rho^2} | \tilde{g}_t]: \text{by (A.35)} \]

By (A.32),

\[ E[\frac{\partial^2 l(\theta_0)}{\partial \rho \partial \tilde{h}} | \tilde{g}_t] = -\frac{1}{\sigma_0^2} E[X_t' W (I_n - \rho_0 W)^{-1} (X_t b_0 + \tilde{\epsilon}_t) | \tilde{g}_t] = -\frac{1}{\sigma_0^2} X_t' C X_t b_0. \] (A.36)

By (A.29) and (A.30),

\[ E[\frac{\partial l(\theta_0)}{\partial \rho} \frac{\partial l(\theta_0)}{\partial \tilde{h}} | \tilde{g}_t] = -\text{tr}(C) E[X_t' \tilde{\epsilon}_t | \tilde{g}_t] + \frac{1}{\sigma_0^4} E[\tilde{\epsilon}_t' W \tilde{g}_t X_t' \tilde{\epsilon}_t | \tilde{g}_t] = \frac{1}{\sigma_0^4} E[X_t' \tilde{\epsilon}_t C X_t b_0 | \tilde{g}_t] + \frac{1}{\sigma_0^4} E[\tilde{\epsilon}_t' C \tilde{\epsilon}_t X_t' \tilde{\epsilon}_t | \tilde{g}_t] = \frac{1}{\sigma_0^4} X_t' C X_t b_0 = -E[\frac{\partial^2 l(\theta_0)}{\partial \rho \partial \tilde{h}} | \tilde{g}_t]: \text{by (A.36)} \]
By (A.32),

\[
E\left[ \frac{\partial^2 l(\theta_0)}{\partial \rho \partial \sigma^2} \mid \tilde{g}_t \right] = -\frac{1}{\sigma_0^4} E[\tilde{e}_t^t C(X_t \tilde{b}_0 + \tilde{e}_t) \mid \tilde{g}_t] = -\frac{1}{\sigma_0^4} E[\tilde{e}_t^t C\tilde{e}_t] = -\frac{1}{\sigma_0^4} \text{tr}(E[\tilde{e}_t^t C])
\]

\[
= -\frac{1}{\sigma_0^2} \text{tr}(C). \tag{A.37}
\]

By (A.29) and (A.30),

\[
E\left[ \frac{\partial l(\theta_0)}{\partial \rho} \frac{\partial l(\theta_0)}{\partial \sigma^2} \mid \tilde{g}_t \right]
\]

\[
= \frac{n \text{tr}(C)}{2\sigma_0^4} - \frac{\text{tr}(C)}{2\sigma_0^4} E[\tilde{e}_t^t \tilde{e}_t] - \frac{n}{2\sigma_0^4} E[\tilde{e}_t^t W \tilde{g}_t \mid \tilde{g}_t] + \frac{1}{2\sigma_0^6} E[\tilde{e}_t^t W \tilde{g}_t \tilde{e}_t^t] \mid \tilde{g}_t]
\]

\[
= -\frac{n}{2\sigma_0^4} E[\tilde{e}_t^t C\tilde{e}_t] + \frac{1}{2\sigma_0^4} E[\tilde{e}_t^t C(X_t \tilde{b}_0 + \tilde{e}_t) \tilde{e}_t^t] \mid \tilde{g}_t] = -\frac{n}{2\sigma_0^2} \text{tr}(C) + \frac{1}{2\sigma_0^6} \text{tr}(CE[\tilde{e}_t^t \tilde{e}_t^t \tilde{e}_t])
\]

\[
= -\frac{n}{2\sigma_0^2} \text{tr}(C) + \frac{1}{2\sigma_0^6} (n + 2)\sigma_0^4 \text{tr}(C) = \frac{1}{\sigma_0^2} \text{tr}(C) = -E\left[ \frac{\partial^2 l(\theta_0)}{\partial \rho \partial \sigma^2} \mid \tilde{g}_t \right], \text{ (by (A.37))} \tag{A.38}
\]

By (A.30),

\[
E\left[ \frac{\partial l(\theta_0)}{\partial b} \frac{\partial l(\theta_0)}{\partial b'} \mid \tilde{g}_t \right] = \frac{1}{\sigma_0^2} E[X_t \tilde{e}_t^t X_t \mid \tilde{g}_t] = \frac{1}{\sigma_0^2} X_t^t X_t = -E\left[ \frac{\partial^2 l(\theta_0)}{\partial b \partial b'} \mid \tilde{g}_t \right],
\]
where the last equality follows from (A.32). By (A.30),

\[
E\left[ \frac{\partial l(\theta_0)}{\partial \sigma^2} \bigg| \tilde{g}_t \right] = -\frac{n}{2\sigma_0^4} E[X_t' \tilde{\epsilon}_t \bigg| \tilde{g}_t] + \frac{1}{2\sigma_0^6} E[\tilde{\epsilon}_t \tilde{\epsilon}_t X_t' \tilde{\epsilon}_t \bigg| \tilde{g}_t]
\]

\[
= -\frac{n}{2\sigma_0^4} X_t' E[\tilde{\epsilon}_t \bigg| \tilde{g}_t] + \frac{1}{2\sigma_0^6} X_t' E[\tilde{\epsilon}_t \tilde{\epsilon}_t \bigg| \tilde{g}_t]
\]

\[
= 0 = -E\left[ \frac{\partial^2 l(\theta_0)}{\partial \sigma^2 \partial b} \bigg| \tilde{g}_t \right],
\]

where the last equality follows from (A.33). At last, by (A.30),

\[
E\left[ \left( \frac{\partial l(\theta_0)}{\partial \sigma^2} \right)^2 \bigg| \tilde{g}_t \right] = E\left[ \left( -\frac{n}{2\sigma_0^4} + \frac{1}{2\sigma_0^6} \tilde{\epsilon}_t \tilde{\epsilon}_t \right)^2 \bigg| \tilde{g}_t \right]
\]

\[
= E\left[ \frac{n^2}{4\sigma_0^4} + \frac{1}{4\sigma_0^6} \tilde{\epsilon}_t \tilde{\epsilon}_t \tilde{\epsilon}_t \tilde{\epsilon}_t - \frac{n}{2\sigma_0^6} \tilde{\epsilon}_t \tilde{\epsilon}_t \right] = \frac{n}{2\sigma_0^4} = -E\left[ \frac{\partial^2 l(\theta_0)}{\partial \sigma^2 \partial \sigma^2} \bigg| \tilde{g}_t \right],
\]

where the last equality follows from (A.33). Hence, we have shown that

\[-E[H(\tilde{y}_t, \tilde{g}_t; \theta_0) \bigg| \tilde{g}_t] = E[s(\tilde{y}_t, \tilde{g}_t; \theta_0)s(\tilde{y}_t, \tilde{g}_t; \theta_0)'] \bigg| \tilde{g}_t] \]

which completes the proof for (vi).

(vii) Suppose for the sake of contradiction that \(E[H(\tilde{y}_t, \tilde{g}_t; \theta_0)]\) is not invertible, then there exists \(a = (a_1, a_2', a_3') \in \mathbb{R}^{2+n(K+1)}\), \(a \neq 0\), such that

\[0 = \alpha' E[H(\tilde{y}_t, \tilde{g}_t; \theta_0)]a = -E[(\alpha' s(\tilde{y}_t, \tilde{g}_t; \theta_0))^2],\]

where the last equality follows from (vi). This implies that \(\alpha' s(\tilde{y}_t, \tilde{g}_t; \theta_0) = 0\), a.s. Denote \(a_2 = (v_1, u_{11}, u_{12}, \ldots, u_{1K}, \ldots, v_n, u_{n1}, u_{n2}, \ldots, u_{nK})'\), \(v = (v_1, \ldots, v_n)'\), \(U = (u_{ij})\). Let
\( C \) be defined in (A.34). Then, we have

\[
0 = a' \bar{y}_t, \bar{g}_t; \theta_0 = -a_1 \text{tr}(C) - \frac{na_3}{2\sigma_0^2} + \frac{a_1}{\sigma_0^2} \bar{\epsilon}_t C \bar{\alpha} + \frac{a_1}{\sigma_0^2} \bar{\epsilon}_t CB_0 \bar{g}_t + \frac{a_1}{\sigma_0^2} \bar{\epsilon}_t \bar{C} \bar{\epsilon}_t \\
+ \frac{1}{\sigma_0^2} v' \bar{\epsilon}_t + \frac{1}{\sigma_0^2} \bar{\epsilon}_t U \bar{g}_t + \frac{a_3}{2\sigma_0^4} \bar{\epsilon}_t, \text{ a.s.} (A.39)
\]

It follows from (A.39) and part (ii) of Assumption 2.5.1 that

\[
0 = -a_1 \text{tr}(C) - \frac{na_3}{2\sigma_0^2} + \frac{a_1}{\sigma_0^2} \bar{\epsilon}' C \bar{\alpha}_0 + \frac{a_1}{\sigma_0^2} \bar{\epsilon}' CB_0 \bar{g}_0 + \frac{a_1}{\sigma_0^2} \bar{\epsilon}' \bar{C} \bar{\epsilon} \\
+ \frac{1}{\sigma_0^2} \bar{\epsilon}' + \frac{1}{\sigma_0^2} \bar{\epsilon}' U \bar{g}_0 + \frac{a_3}{2\sigma_0^2} \bar{\epsilon}' \bar{\epsilon}, \text{ for any } (\bar{\epsilon}', \bar{g}')(\bar{\epsilon}', \bar{g}')(A.40)
\]

By taking partial derivatives with respect to \((\bar{\epsilon}', \bar{g}')(A.40)) on both sides of (A.40), we obtain that

\[
-a_1 \text{tr}(C) - \frac{na_3}{2\sigma_0^2} = 0, \frac{a_1}{\sigma_0^2} \bar{\epsilon}' C \bar{\alpha}_0 + \frac{1}{\sigma_0^2} v = 0, \frac{a_1}{\sigma_0^2} CB_0 + \frac{1}{\sigma_0^2} U = 0, (A.41)
\]

\[
\frac{a_1}{\sigma_0^2} C + \frac{a_3}{2\sigma_0^4} I_n = 0. (A.42)
\]

There are two cases:

Case 1: \( a_1 = 0 \). Then, it follows from (A.41) that \( a_3 = 0, v = 0, \) and \( U = 0 \), which contradict to \( a \neq 0 \).

Case 2: \( a_1 \neq 0 \). Then, it follows from (A.42) that \( C = \frac{a_3}{2\sigma_0^2 a_1} I_n \), which in combination with (A.34) implies that \( W(1 - \frac{a_3 \rho_0}{2\sigma_0^2 a_1}) = \frac{a_3}{2\sigma_0^2 a_1} I_n \). If \( a_3 = 0 \), then \( W = 0 \), which contradicts to Assumption 2.5.2; if \( a_3 \neq 0 \), then \( 1 - \frac{a_3 \rho_0}{2\sigma_0^2 a_1} \neq 0, \) and \( W = \frac{a_3}{2\sigma_0^2 a_1 - a_3 \rho_0} I_n \),
which contradicts to that the diagonal elements of \( W \) are zero (Assumption 2.5.2). Hence, \( E[H(\hat{y}_t, \tilde{g}_t; \theta_0)] \) is invertible.

(viii) Let \( \mathcal{N} \) be any neighborhood of \( \theta_0 \) that lies in the interior of \( \Theta \). We have

\[
E[\sup_{\theta \in \mathcal{N}} \| H(\hat{y}_t, \tilde{g}_t; \theta) \|] \leq E[\sup_{\theta \in \mathcal{N}} \left| \frac{\partial^2 l(\theta)}{\partial \rho^2} \right|] + E[\sup_{\theta \in \mathcal{N}} \left| \frac{\partial^2 l(\theta)}{\partial \rho \partial \sigma_2} \right|] + E[\sup_{\theta \in \mathcal{N}} \left| \frac{\partial^2 l(\theta)}{\partial \rho \partial \sigma} \right|] + E[\sup_{\theta \in \mathcal{N}} \left| \frac{\partial^2 l(\theta)}{\partial \sigma^2 \partial \sigma_2} \right|] + E[\sup_{\theta \in \mathcal{N}} \left| \frac{\partial^2 l(\theta)}{\partial \sigma \partial \sigma} \right|].
\]

(A.43)

We only need to show that each term on the right side of (A.43) is finite. Let \( C(\rho) := W(I_n - \rho W)^{-1} \), then \( C(\rho) \) is continuous. By (A.31), (A.32), and (A.33), we have

\[
E[\sup_{\theta \in \mathcal{N}} \left| \frac{\partial^2 l(\theta)}{\partial \rho^2} \right|] \leq \sup_{\theta \in \mathcal{N}} \left| \text{tr}(C(\rho)^2) \right| + E[\| \tilde{y}_t W' W \tilde{y}_t \|] \sup_{\theta \in \mathcal{N}} \frac{1}{\sigma^2},
\]

(A.44)

\[
E[\sup_{\theta \in \mathcal{N}} \| \frac{\partial^2 l(\theta)}{\partial \rho \partial \delta} \|] \leq E[\| X_t' W \tilde{y}_t \|] \sup_{\theta \in \mathcal{N}} \frac{1}{\sigma},
\]

(A.45)

\[
E[\sup_{\theta \in \mathcal{N}} \left| \frac{\partial^2 l(\theta)}{\partial \sigma^2 \partial \sigma_2} \right|] = E[\sup_{\theta \in \mathcal{N}} \left| \frac{1}{\sigma^2} \tilde{\zeta}_t W \tilde{y}_t \right|]
\]

\[
\leq E[\| \tilde{y}_t W \tilde{y}_t \|] \sup_{\theta \in \mathcal{N}} \frac{1}{\sigma^4} + E[\| \tilde{y}_t W' W \tilde{y}_t \|] \sup_{\theta \in \mathcal{N}} \frac{\rho}{\sigma^4} + E[\| W \tilde{y}_t \|] \sup_{\theta \in \mathcal{N}} \frac{1}{\sigma} \| \tilde{\alpha} \|
\]

\[
+ E[\| \tilde{g}_t \|]^2 \sup_{\theta \in \mathcal{N}} \frac{1}{2\sigma^4} \| B \|^2 + E[\| W \tilde{g}_t \|]^2 \sup_{\theta \in \mathcal{N}} \frac{1}{2\sigma^4},
\]

(A.46)

\[
E[\sup_{\theta \in \mathcal{N}} \left| \frac{\partial^2 l(\theta)}{\partial \delta \partial \delta} \right|] \leq E[\| X_t' X_t \|] \sup_{\theta \in \mathcal{N}} \frac{1}{\sigma^2},
\]

(A.47)

\[
E[\sup_{\theta \in \mathcal{N}} \left| \frac{\partial^2 l(\theta)}{\partial \sigma^2 \partial \sigma} \right|] = E[\sup_{\theta \in \mathcal{N}} \left| \frac{1}{\sigma^4} X_t' \tilde{\xi}_t \right|]
\]

99
\[ \leq E[\sup_{\theta \in \mathcal{N}}\left| \frac{1}{\sigma^2}X'\tilde{y}_t \right|] + E[\sup_{\theta \in \mathcal{N}}\frac{\rho}{\sigma^2}X'W\tilde{y}_t] + E[\sup_{\theta \in \mathcal{N}}\left| \frac{1}{\sigma^2}X'\tilde{\alpha} \right|] + E[\sup_{\theta \in \mathcal{N}}\left| \frac{1}{\sigma^2}X'B\tilde{g}_t \right|] \]
\[ \leq E[\|X'\tilde{y}_t\|] \sup_{\theta \in \mathcal{N}}\frac{1}{\sigma^4} + E[\|X'W\tilde{y}_t\|] \sup_{\theta \in \mathcal{N}}\frac{\rho}{\sigma^4} + E[\|X_t\|] \sup_{\theta \in \mathcal{N}}\frac{\|\tilde{\alpha}\|}{\sigma^4} + E[\|X'_t\|\|\tilde{g}_t\|] \sup_{\theta \in \mathcal{N}}\frac{\|B\|}{\sigma^4} \]
\[ \leq E[\|X'\tilde{y}_t\|] \sup_{\theta \in \mathcal{N}}\frac{1}{\sigma^4} + E[\|X'W\tilde{y}_t\|] \sup_{\theta \in \mathcal{N}}\frac{\rho}{\sigma^4} + E[\|X_t\|] \sup_{\theta \in \mathcal{N}}\frac{\|\tilde{\alpha}\|}{\sigma^4} + E[\|X'_t\|\|\tilde{g}_t\|] \sup_{\theta \in \mathcal{N}}\frac{\|B\|}{\sigma^4} \]
\[ + \frac{1}{2}(E[\|X'_t\|^2] + E[\|\tilde{g}_t\|^2]) \sup_{\theta \in \mathcal{N}}\frac{\|B\|}{\sigma^4}, \quad (A.48) \]
\[ E[\sup_{\theta \in \mathcal{N}}\left| \frac{\partial^2 l(\theta)}{\partial \sigma^2 \partial \sigma^2} \right|] = E[\sup_{\theta \in \mathcal{N}}\left| \frac{n}{2\sigma^4} - \frac{1}{\sigma^6} \tilde{\xi}'t\tilde{\xi}_t \right|] \]
\[ = E[\sup_{\theta \in \mathcal{N}}\left| \frac{n}{2\sigma^4} - \frac{1}{\sigma^6} \|(I_n - \rho W)\tilde{y}_t - \tilde{\alpha} - B\tilde{g}_t\|^2 \right|] \]
\[ \leq E[\sup_{\theta \in \mathcal{N}}\left( \frac{n}{2\sigma^4} + \frac{16}{\sigma^6}(\|\tilde{y}_t\|^2 + \rho^2\|W\tilde{y}_t\|^2 + \|\tilde{\alpha}\|^2 + \|B\|^2\|\tilde{g}_t\|^2) \right)] \]
\[ \leq \sup_{\theta \in \mathcal{N}}(\frac{n}{2\sigma^4} + \frac{16\|\tilde{\alpha}\|^2}{\sigma^6}) + E[\|\tilde{y}_t\|^2] \sup_{\theta \in \mathcal{N}}\frac{16}{\sigma^6} + E[\|W\tilde{y}_t\|^2] \sup_{\theta \in \mathcal{N}}\frac{16\rho^2}{\sigma^6} + E[\|\tilde{g}_t\|^2] \sup_{\theta \in \mathcal{N}}\frac{16\|B\|^2}{\sigma^6}. \quad (A.49) \]

Since $\Theta$ is compact, all the supremums on the right-hand side of (A.44)-(A.49) are finite.

Furthermore, by part (i) of Assumption 2.5.1, $\tilde{g}_t$ and hence $\tilde{y}_t$ have finite second moments.

Thus, all the expectations on the right-hand side of (A.44)-(A.49) are finite. Therefore, each term on the right-hand side of (A.43) is finite, which completes the proof. \qed
Appendix B

Appendix for Chapter 3

This appendix mainly concerns the MCMC Subroutines and Proofs for the SV-DEJ model. The MCMC procedures for the SV-MJ-JV model and the SV-VG model have been provided in Eraker, Johannes and Polson (2003) and Li, Wells and Yu (2008). Thus in what follows, we focus on the MCMC procedures for the SV-DEJ model. We mostly choose conjugate priors to ease computation. The priors are relatively not informative and are used in previous study; see the values of hyper-parameters below and also the prior mean and standard deviation in Table 3.2. We also check the robustness of the results reported in Section 3.4 by specifying different combinations of hyper-parameters. Results do not seem to differ significantly. Bayesian MCMC inferences are done on the parameter space $\Theta$ (say, under the SV-DEJ model, $\Theta = \{\mu, \kappa, \theta, \rho, \sigma_v, \eta^+, \eta^-, \lambda^+, \lambda^-\}$), latent variables $N_{1:T}$ (the
jump times), $V_{0:T}$ (the volatilities), $\xi_{1:T}^+$ (positive jump sizes) and $\xi_{1:T}^-$ (negative jump sizes),
given the returns data $Y_{0:T}$. Below are details of all conditional posteriors.

First, we introduce the conditional posteriors for parameters and latent variables of the
SV-DEJ model that are shared with the SV-MJ-JV model. One can easily discover the
similarity by comparing the results here and those in A.2 of Li, Wells and Yu (2008). Thus
the proofs are omitted for the following conditional posteriors.

1. **Posterior for $\mu$** The posterior of $\mu$ follows a normal distribution $\mu \sim N\left(\frac{S^{(\mu)}}{W^{(\mu)}},\frac{1}{W^{(\mu)}}\right)$,
   where
   \[
   W^{(\mu)} = \left(\frac{1}{1-\rho^2}\right) \left(\sum_{t=0}^{T-1} \frac{1}{v_t}\right) + \frac{1}{M^{(\mu)^2}},
   S^{(\mu)} = \frac{1}{(1-\rho^2)} \sum_{t=0}^{T-1} \frac{1}{v_t} \left( C^{(\mu)}_{t+1} - \rho D^{(\mu)}_{t+1}\sigma_v \right),
   m^{(\mu)} = 0, M^{(\mu)} = 1
   \]

2. **Posterior for $\theta$** The posterior of $\theta$ follows a truncated normal distribution $\theta \sim N\left(\frac{S^{(\theta)}}{W^{(\theta)}},\frac{1}{W^{(\theta)}}\right) 1_{(\theta>0)}$, where
   \[
   W^{(\theta)} = \frac{\kappa^2 \Delta}{\sigma_v^2(1-\rho^2)} \left(\sum_{t=0}^{T-1} \frac{1}{v_t}\right) + \frac{1}{M^{(\theta)^2}},
   S^{(\theta)} = \frac{\kappa}{(1-\rho^2)} \sum_{t=0}^{T-1} \left( \frac{D^{(\theta)}_{t+1}/\sigma_v - \rho C^{(\theta)}_{t+1}}{v_t} \right) + \frac{m^{(\theta)}}{M^{(\theta)^2}},
   C^{(\theta)}_{t+1} = y_{t+1} - y_t - \mu \Delta - J^y_{t+1},
   D^{(\theta)}_{t+1} = v_{t+1} + (\kappa \Delta - 1)v_t.
   \]
   Here the prior for $\theta$ is $N(m^{(\theta)}, M^{(\theta)^2}) 1_{(\theta>0)}$, where $m^{(\theta)} = 0, M^{(\theta)} = 1$ in our study.

3. **Posterior for $\kappa$** The posterior of $\kappa$ follows a truncated normal distribution $\kappa \sim
\[
N \left( \frac{S^{(\kappa)}}{W^{(\kappa)}}, \frac{1}{W^{(\kappa)}} \right) 1_{(\kappa>0)}, \quad \text{where} \quad W^{(\kappa)} = \frac{\Delta}{\sigma_v^2(1-\rho^2)} \left( \sum_{t=0}^{T-1} \frac{(\theta-v_t)^2}{\nu_t} \right) + \frac{1}{M^{(\kappa)}}, \quad S^{(\kappa)} = \\
\frac{1}{(1-\rho^2)\sigma_v} \sum_{t=0}^{T-1} \frac{(\theta-v_t)(D_t^{(\kappa)}/\sigma_v - \rho C_t^{(\kappa)})}{\nu_t} + \frac{m^{(\kappa)}}{M^{(\kappa)}}, \quad C_{t+1} = y_{t+1} - y_t - \mu \Delta - J_y^{(t+1)}, \quad D_{t+1} = \nu_{t+1} - \nu_t.
\]

Here the prior for \( \kappa \) is \( N(m^{(\kappa)}, M^{(\kappa)}^2)1_{(\kappa>0)} \), where \( m^{(\kappa)} = 0 \), \( M^{(\kappa)} = 1 \) in our study.

4. **Posterior for \( \sigma_v \) and \( \rho \)** Following Jacquier, Polson and Rossi (1994), we transform \((\rho, \sigma_v)\) one-to-one into \((\phi_v, w_v)\), where \( \phi_v = \sigma_v \rho \) and \( w_v = \sigma_v^2(1-\rho^2) \). The priors are a normal-inverse-gamma distribution: \( \phi_v|w_v \sim N(0, 1/w_v) \) and \( w_v \sim \text{IG}(m^{(RS)}, M^{(RS)}) \), where \( m^{(RS)} = 2 \), \( M^{(RS)} = 200 \) in our study. Given this reparameterization, the joint posteriors of \((\phi_v, w_v)\) are the conjugate of the priors

\[
w_v \sim \text{IG} \left( \frac{T}{2} + m^{(RS)}, \frac{1}{\frac{1}{2} \sum_{t=0}^{T-1} (D_t^{(RS)})^2 + \frac{1}{M^{(RS)}} - \frac{S^{(RS)}^2}{2W^{(RS)}}} \right)
\]

and

\[
\phi_v|w_v \sim N \left( \frac{S^{(RS)}}{W^{(RS)}}, \frac{w_v}{W^{(RS)}} \right),
\]

where \( W^{(RS)} = \sum_{t=0}^{T-1} (C_t^{(RS)})^2 + 2 \), \( S^{(RS)} = \sum_{t=0}^{T-1} C_t^{(RS)} D_t^{(RS)} \), \( C_t^{(RS)} = (y_{t+1} - y_t - \mu \Delta - J_y^{(t+1)})/\sqrt{\nu_t \Delta} \), \( D_{t+1}^{(RS)} = (\nu_{t+1} - \nu_t - \kappa(\theta - \nu_t)\Delta)/\sqrt{\nu_t \Delta} \).
5. **Posterior for the latent variable** \( v_{t+1} \)  
For \( 0 < t + 1 < T \), the posterior of \( v_{t+1} \) is

\[
p(v_{t+1} | \cdot) \propto \exp \left[ - \frac{2 \rho \epsilon^v_{t+1} + (\epsilon^v_{t+1})^2}{2(1 - \rho^2)} \right] \times \frac{1}{v_{t+1}} \times \exp \left[ - \frac{(\epsilon^y_{t+2})^2 - 2 \rho \epsilon^y_{t+2} \epsilon^v_{t+2} + (\epsilon^v_{t+2})^2}{2(1 - \rho^2)} \right]
\]

where

\[
\begin{align*}
\epsilon^y_{t+1} &= \frac{(y_{t+1} - y_t - \mu \Delta - J^y_{t+1})}{\sqrt{v_t \Delta}}, \\
\epsilon^v_{t+1} &= \frac{(v_{t+1} - v_t - \kappa(\theta - v_t) \Delta)}{(\sigma_v \sqrt{v_t \Delta})}.
\end{align*}
\]

For \( t + 1 = T \), the above posterior only has the first exponential part because \( v_T \) depends only on \( v_{T-1} \). Similarly, the posterior of \( p(v_0 | \cdot) \) depends on \( \frac{1}{v_0} \) and the second exponential part. Following Li, Wells and Yu (2008), we make use of the adaptive rejection metropolitan sampling method of Gilks, Best and Tan (1994) to draw from this conditional posterior.

Next, we turn to parameters and variables that are special to SV-DEJ. The conditional posteriors are new and they are derived as follows.

1. **Posterior for** \( \eta^+, \eta^- \)  
The posterior of \( \eta^+ \) follows an inverse-gamma distribution

\[
\eta^+ \sim IG \left( T + m^{(EP)}, \frac{1}{\sum_{t=0}^{T-1} s_t + 1/M^{(EP)}} \right),
\]

\[
\eta^- \sim IG \left( T - m^{(EP)}, \frac{1}{\sum_{t=0}^{T-1} s_t + 1/M^{(EP)}} \right).
\]
where the prior of $\eta^+$ is $IG(m^{(EP)}, M^{(EP)})$ and $m^{(EP)} = 20, M^{(EP)} = 10$ in our study. Similarly, the posterior of $\eta^-$ also follows an inverse-gamma distribution

$$\eta^- \sim IG \left( T + m^{(EM)}, \frac{1}{\sum_{t=0}^{T-1} \xi_{t+1}^- + 1/M^{(EM)}} \right),$$

where the prior of $\eta^-$ is $IG(m^{(EM)}, M^{(EM)})$ and $m^{(EM)} = 20, M^{(EM)} = 10$ in our study.

**Proof.** Since the two cases for $\eta^+$ and $\eta^-$ are almost the same, we focus on the case of $\eta^+$ only. To ease notation, in this proof, we denote $m = m^{(EP)}$ and $M = M^{(EP)}$.

By Bayes’ rule,

\[
p(\eta^+|\Theta \setminus \{\eta^+\}, N_{1:T}, Y_{0:T}, V_{0:T}, \xi^+_{1:T}, \xi^-_{1:T}) = p(\xi^+_{1:T} | \eta^+) \propto p(\xi^+_{1:T} | \eta^+) p(\eta^+)p(\eta^+)
\]

\[
\propto \prod_{t=0}^{T-1} \left[ (\eta^+)^{-1} \exp\left(-\xi^+_{t+1}/\eta^+\right) \right] \cdot (\eta^+)^{-m-1} \exp\left(-\frac{1}{M^{(EM)}}\right)
\]

\[
\propto (\eta^+)^{-(m+T)-1} \exp\left(-\sum_{t=0}^{T-1} \xi^-_{t+1} + 1/M^{(EM)}\right)
\]

where the equality results from independency. This is readily recognized as a standard exponential-gamma conjugate pair model. \qed

2. **Posterior for the latent variables** $\xi^+_{t+1}, \xi^-_{t+1}$ For $1 \leq t + 1 \leq T$, the posterior of
\( \xi_{t+1}^+ \) follows a truncated normal distribution when \( N_{t+1} = 1 \)

\[
\xi_{t+1}^+ \sim N \left( \frac{S^{(XP)}}{W^{(XP)}}, \frac{1}{W^{(XP)}} \right) 1_{(\xi_{t+1}^+ > 0)},
\]

where \( S^{(XP)} = \frac{C^{(XP)}_{t+1} - \rho D^{(XP)}_{t+1}}{\sigma_v} - \frac{1}{\eta^+}, \quad W^{(XP)} = \frac{1}{\eta^+}, \quad C^{(XP)}_{t+1} = y_{t+1} - y_t - \mu \Delta \)
and \( D^{(XP)}_{t+1} = v_{t+1} - v_t - \kappa (\theta - v_t) \Delta \). When \( N_{t+1} \neq 1 \), the posterior is the same as its prior \( \xi_{t+1}^+ \sim \exp (\eta^+) \).

Similarly, for \( 1 \leq t + 1 \leq T \), the posterior of \( \xi_{t+1}^- \) follows a truncated normal distribution when \( N_{t+1} = -1 \)

\[
\xi_{t+1}^- \sim N \left( \frac{S^{(XM)}}{W^{(XM)}}, \frac{1}{W^{(XM)}} \right) 1_{(\xi_{t+1}^- > 0)},
\]

where \( S^{(XM)} = -\frac{C^{(XM)}_{t+1} - \rho D^{(XM)}_{t+1}}{\sigma_v} - \frac{1}{\eta^-}, \quad W^{(XM)} = \frac{1}{\eta^-} \). And when \( N_{t+1} \neq -1 \), the posterior is the same as its prior \( \xi_{t+1}^- \sim \exp (\eta^-) \).

**Proof.** Since the cases for \( N_{t+1} = 1 \) and \( N_{t+1} = -1 \) are similar, we focus on the case where \( N_{t+1} = 1 \). To ease notation, in this proof, we denote \( m = m^{(XP)} \), \( M = M^{(XP)} \), \( C_{t+1} = C^{(XP)}_{t+1} \), \( D_{t+1} = D^{(XP)}_{t+1} \), \( S = S^{(XP)} \) and \( W = W^{(XP)} \). It is easy to see the posterior is \( \xi_{t+1}^+ \sim \exp (\eta^+) \) when \( N_{t+1} \neq 1 \) since the data provides
no information. By Bayes’ rule, when $N_{t+1} = 1$,

\[
p(\xi_{t+1}^+ | \cdot) = p(\xi_{t+1}^+ | y_{t+1}, y_t, v_{t+1}, v_t, N_{t+1} = 1, \Theta) \\
\propto p(y_{t+1} | y_t, \Theta, N_{t+1} = 1, v_t, v_{t+1}, \xi_{t+1}^+) p(\xi_{t+1}^+ | \eta^+) \\
\propto \exp \left[ -\frac{1}{2v_t} \Delta (1 - \rho^2) (C_{t+1} - \xi_{t+1}^+ - \rho D_{t+1}/\sigma_v)^2 \right] \exp \left( -\frac{\xi_{t+1}^+}{\eta^+} \right) 1_{(\xi_{t+1}^+ > 0)} \\
\propto \exp \left( -\frac{1}{2} W(\xi_{t+1}^+)^2 - S_{t+1} \xi_{t+1}^+ \right) 1_{(\xi_{t+1}^+ > 0)},
\]

The required conclusion readily follows by completing the squares and comparing the parameters. 

3. **Posterior for $(\lambda^+, \lambda_0, \lambda^-)$** The posterior of $(\lambda^+, \lambda_0, \lambda^-)$ follows a Dirichlet distribution

\[
p((\lambda^+, \lambda_0, \lambda^-) | \cdot) \\
\sim D \left( \alpha_1 + \sum_{t=0}^{T-1} 1_{(N_{t+1}=1)}, \alpha_0 + \sum_{t=0}^{T-1} 1_{(N_{t+1}=0)}, \alpha_{-1} + \sum_{t=0}^{T-1} 1_{(N_{t+1}=-1)} \right),
\]

where the prior of $(\lambda^+, \lambda_0, \lambda^-)$ is $D(\alpha_1, \alpha_0, \alpha_{-1})$ and $\alpha_1 = \alpha_{-1} = 2$ and $\alpha_0 = 40$ in our study.
Proof. By Bayes’ rule,

\[ p(\lambda^+, \lambda_0, \lambda^-|\cdot) = p(\lambda^+, \lambda_0, \lambda^-|N_{1:T}) \]

\[ \propto p(N_{1:T}|(\lambda^+, \lambda_0, \lambda^-)) p(\lambda^+, \lambda_0, \lambda^-) \]

\[ \propto (\lambda^+)^{\sum_{t=0}^{T-1} 1(N_{t+1} = 1)} (\lambda^-)^{\sum_{t=0}^{T-1} 1(N_{t+1} = 0)} (\lambda^-)^{\sum_{t=0}^{T-1} 1(N_{t+1} = -1)} \]

\[ = (\lambda^+)^{\sum_{t=0}^{T-1} 1(N_{t+1} = 1) + \alpha_1 - 1} (\lambda^-)^{\sum_{t=0}^{T-1} 1(N_{t+1} = 0) + \alpha_0 - 1} \]

\[ \sim D \left( \alpha_1 + \sum_{t=0}^{T-1} 1(N_{t+1} = 1), \alpha_0 + \sum_{t=0}^{T-1} 1(N_{t+1} = 0), \alpha_{-1} + \sum_{t=0}^{T-1} 1(N_{t+1} = -1) \right). \]

\[ \square \]

4. Posterior for the latent variable \( N_{t+1} \)  For \( 1 \leq t + 1 \leq T \), the posterior of \( N_{t+1} \)
follows a trinomial distribution

\[ p(N_{t+1} = i|\cdot) = \begin{cases} 
\frac{\lambda^+ \exp(U_i)}{S(N)}, & \text{when } i = 1; \\
\frac{\lambda_0 \exp(U_0)}{S(N)}, & \text{when } i = 0; \\
\frac{\lambda^- \exp(U_{-1})}{S(N)}, & \text{when } i = -1,
\end{cases} \]
where

\[ U_1 = -\frac{1}{2v_t \Delta (1 - \rho^2)} \left[ (C_{t+1}^{(N)} - \xi_{t+1}^+)^2 - 2\rho (C_{t+1}^{(N)} - \xi_{t+1}^+) D_{t+1}^{(N)} / \sigma_v \right], \]

\[ U_0 = -\frac{1}{2v_t \Delta (1 - \rho^2)} \left[ (C_{t+1}^{(N)})^2 - 2\rho C_{t+1}^{(N)} D_{t+1}^{(N)} / \sigma_v \right], \]

\[ U_{-1} = -\frac{1}{2v_t \Delta (1 - \rho^2)} \left[ (C_{t+1}^{(N)} + \xi_{t+1}^-)^2 - 2\rho (C_{t+1}^{(N)} + \xi_{t+1}^-) D_{t+1}^{(N)} / \sigma_v \right], \]

\[ S^{(N)} = p_1 \exp(U_1) + \lambda_0 \exp(U_0) + \lambda^- \exp(U_{-1}), \]

\[ C_{t+1}^{(N)} = y_{t+1} - y_t - \mu \Delta, \]

\[ D_{t+1}^{(N)} = v_{t+1} - v_t - \kappa (\theta - v_t) \Delta. \]

**Proof.** To ease notations, in this proof, define \( p_1 := \lambda^+, \lambda_0 := \lambda_0 \) and \( p_{-1} := \lambda^- \). By Bayes’ rule, we have for \( i = -1, 0, 1, \)

\[ p(N_{t+1} = i|\cdot) = p(N_{t+1} = i|\Theta, y_t, y_{t+1}, v_t, v_{t+1}, \xi_{t+1}^+, \xi_{t+1}^-) \]

\[ \quad = \frac{p(y_{t+1}|y_t, \Theta, N_{t+1} = i, v_t, v_{t+1}, \xi_{t+1}^+, \xi_{t+1}^-)p(N_{t+1} = i|\Theta)}{\sum_j p(y_{t+1}|y_t, \Theta, N_{t+1} = j, v_t, v_{t+1}, \xi_{t+1}^+, \xi_{t+1}^-)p(N_{t+1} = j|\Theta)} \]

\[ = \frac{p(y_{t+1}|y_t, \Theta, N_{t+1} = i, v_t, v_{t+1}, \xi_{t+1}^+, \xi_{t+1}^-)p_i}{\sum_j p(y_{t+1}|y_t, \Theta, N_{t+1} = j, v_t, v_{t+1}, \xi_{t+1}^+, \xi_{t+1}^-)p_j}. \]

Since all three cases are similar, only the case of \( i = 1 \) is considered here. When
\( i = 1, \)

\[
p(y_{t+1}|y_t, \Theta, N_{t+1} = i, v_t, v_{t+1}, \xi_{t+1}, \xi_{t+1}^-) = p(y_{t+1}|y_t, \Theta, N_{t+1} = 1, v_t, v_{t+1}, \xi_{t+1}^+) \\
= \frac{1}{\sqrt{2\pi v_t \Delta (1 - \rho^2)}} \exp \left[ -\frac{1}{2v_t \Delta (1 - \rho^2)} \left( C_{t+1} - \xi_{t+1}^+ - \rho D_{t+1}/\sigma_v \right)^2 \right]
\]

The required conclusion readily follows when all common factors are canceled out.

\[ \square \]
Bibliography


Duffie, D. and Pan, J. and Singleton, K., 2000. Transform analysis and asset pricing for


and Expected Financial Asset Risk and Return: Theory and Empirical Evidence. The

Fernandez, V., 2011. Spatial Linkages in International Financial Markets. Quantitative Fi-
nance 11, 237–245.

American Statistical Association 56, 971–977.

Gautschi W., 2006. Computing the Kantorovich-Lebedev Integral Transforms and Their


