Asset Pricing with Spatial Interaction*

Steven Kou† Xianhua Peng‡ Haowen Zhong§

Abstract

We propose a spatial capital asset pricing model (S-CAPM) and a spatial arbitrage pricing theory (S-APT) that extend the classical asset pricing models by incorporating spatial interaction. We then apply the S-APT to study the comovements of Eurozone stock indices (by extending the Fama-French factor model to regional stock indices) and the futures contracts on S&P/Case-Shiller Home Price Indices; in both cases spatial interaction is significant and plays an important role in explaining cross-sectional correlation.

Keywords: capital asset pricing model, arbitrage pricing theory, spatial interaction, real estate, factor model, property derivatives, futures

JEL classification: G12, G13, R31, R32

1 Introduction

A central issue in financial economics is to understand the risk-return relationship for financial assets, as exemplified by the classical capital asset pricing model (CAPM)

*We are grateful to the seminar and conference participants at Ajou University, Chinese University of Hong Kong, Columbia University, Hong Kong University of Science and Technology, National University of Singapore, University of California at Berkeley, University of Texas at Austin, INFORMS Annual Conferences (2011, 2012), American Real Estate Society 2012 Annual Meeting, Pacific Rim Real Estate Society 2012 Annual Conference, Asian Real Estate Society and American Real Estate and Urban Economics Association Joint 2012 Conference, Asian Quantitative Finance Conference 2013, as well as the Institute of Mathematical Statistics Annual Meeting 2014 for their helpful comments and discussion. This research is supported by the National Science Foundation of the United States, the University Grants Committee of HKSAR of China (Project No. FS-GRF14SC19), the School of Science (Project No. R9307) and the Department of Mathematics of HKUST.

†Risk Management Institute and Department of Mathematics, National University of Singapore, 21 Heng Mui Keng Terrace, I3 Building 04-03, Singapore 119613. Email: matsteve@nus.edu.sg.

‡Corresponding author. Department of Mathematics, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong. Email: maxhpeng@ust.hk. Tel: (852)23587431. Fax: (852)23581643.

§Department of Industrial Engineering & Operations Research, Columbia University, 500 West 120th Street, New York, NY 10021. Email: hz2193@columbia.edu.

and arbitrage pricing theory (APT). Building on the seminal work of Markowitz (1952), CAPM, as proposed by Sharpe (1964) and others, characterizes the market equilibrium when all market participants hold mean-variance efficient portfolios. Unlike the CAPM, the APT introduced by Ross (1976a,b) is based on an asymptotic arbitrage argument rather than on market equilibrium, which allows for multiple risk factors and does not require the identification of the market portfolio; see e.g., Huberman (1982), Chamberlain (1983), Chamberlain and Rothschild (1983), Ingersoll (1984), Huberman and Wang (2008), among others.

In terms of empirical performance, APT improves on CAPM in that cross-sectional differences in expected asset returns are better accounted for by multiple factors in APT; see the Nobel-prize-winning work of Fama and French (1993) and Fama and French (2012), among others. Going beyond expected returns, however, the APT models with the famous factors in existing literature do not seem to capture all the cross-sectional variations in realized asset returns. In particular, a motivating example of fitting an APT model to the European countries stock indices returns (see Section 2.1) shows that (i) there is evidence of cross-sectional spatial interaction among the residuals of the APT regression model; (ii) the no asymptotic arbitrage constraint (i.e., zero-intercept constraint) implied by the APT is rejected by the data, which demonstrates that the existing APT model and famous factors are not adequate in accounting for the cross-sectional variatioins.

To better account for potential spatial correlation among residuals of APT models and to better capture the no asymptotic arbitrage constraint, in this paper, we attempt to link spatial econometrics, which emphasizes the statistical modeling of spatial interaction, with the classical CAPM and APT. Empirical importance of spatial interaction has already been found in the real estate markets (see, e.g., Anselin (1988) and Cressie (1993)), in U.S. equity market (see Pirinsky and Wang (2006) for comovements of common stock returns of US corporations in the same geographic area), and in international stock portfolios (see Bekaert, Hodrick, and Zhang (2009)). Coval and Moskowitz (2001) demonstrated empirically the importance of spatial information in the investment decisions and outcomes of individual fund managers.

In this paper we study the impact of spatial information on overall markets in the form of CAPM or APT. More precisely, we first propose a spatial capital asset pricing model (S-CAPM) and a spatial arbitrage pricing theory (S-APT), and then study empirical implications of the models. The new models can be applied to financial assets that can be sold short, such as national/regional stock indices and futures contracts on the S&P/Case-Shiller Home Price Indices (Case and Shiller (1987)).

Our S-CAPM and S-APT differ from existing models in spatial econometrics. The consideration of equilibrium pricing and no arbitrage pricing imposes certain constraints on the parameters in the S-CAPM and S-APT models (see Eq. (15) and Eq. (26) in Theorem 4.1); 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 generally not subject to constraints.

After developing the economic models, we give two applications of the proposed S-APT. First, we continue the investigation of the motivating example in Section 2.1, in which the comovements of returns of Eurozone stock indices are studied, by extending the factor model for international stocks proposed in Fama and French (2012) to incorporate spatial interaction. Factor models for stock markets have been well studied in the literature. In the ground-breaking work of Fama and French (1993), two factors related to firm size and book-to-market equity are constructed and shown to have great explanatory power of cross-sectional stock returns. In their approach, a factor is constructed as the difference between the returns of firms with certain characteristics (e.g. small-cap) and those with opposite characteristics (e.g. large-cap). This approach has at least two advantages. First, factors constructed in this way are payoffs of zero-cost portfolios that are traded in the market and further steps of linear projection of factors are unnecessary. Second, the factors have clear economic interpretation. For instance, since the book-to-market ratio is indicative of financial distress, the factor constructed according to the ratio can be viewed as a proxy for distress risk. Fama and French (2012) investigate the performance of the market, size, value, and momentum factors in international stock markets.

We extend Fama and French (2012) in three ways: (i) By using the S-APT model instead of APT models (factor models without spatial interaction), we investigate the role of spatial interaction in the explanation of comovements of stock index returns. We find that spatial interaction is significant, even after controlling for popular factors, including market, size, value, and momentum factor. (ii) Adding spatial interaction for the comovements not only improves the overall model fitting in terms of Akaike information criterion (AIC), but also reduces the degree of spatial correlation (i.e., κ in Eq. (2)) among residuals of the fitted model. Furthermore, the no asymptotic arbitrage constraint (i.e., the zero-intercept constraint) is no longer rejected by the data after spatial interaction is incorporated in the model. (iii) We focus on Eurozone stock indices that are portfolios of stocks implicitly sorted by locations/nations, while Fama and French (2012) study the returns of stock portfolios constructed according to other issuer characteristics such as size and value.

As the second application, we apply the S-APT model to the study of risk-return relationship of real estate securities, particularly the S&P/Case-Shiller Home Price Indices (CSI Indices) futures. 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. 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; see Fabozzi, Shiller, and Tunaru (2012) and the references therein for the pricing and use of property derivatives for risk management.

We add to the literature on the study of real estate securities by constructing a three-factor S-APT model for the CSI Indices futures returns. Using monthly return data, we find that the spatial interaction among CSI indices futures returns are significantly positive. In addition, the no asymptotic arbitrage (i.e., zero-intercept) constraint implied by the S-APT model is not rejected by the futures data. Furthermore, incorporating spatial interaction improves the model fitting to the data in terms of AIC and eliminates the spatial correlation among the residuals of APT models.

In summary, the main contribution of this paper is twofold: (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. In addition, we develop estimation and testing procedures for implementing the S-APT model. (ii) Empirically, we apply the S-APT models to the study of the Eurozone stock indices returns and the futures contracts written on the CSI Indices. In both cases, the spatial interaction incorporated in the S-APT model seems to be a significant factor in explaining asset return comovements.

The remainder of the paper is organized as follows. In Section 2, a motivating example is discussed and a linear model with spatial interaction is introduced. The S-CAPM and S-APT for ordinary assets and futures contracts are derived in Sections 3 and 4, respectively. Section 5 develops the econometric tools for implementing the S-APT model. The rigorous econometric analysis of the identification and statistical inference problems for the proposed spatial econometric model is given in Appendix D. The empirical studies on the Eurozone stock indices and the CSI Indices futures using the S-APT are provided in Sections 6 and 7, respectively. Section 8 concludes.

2 Preliminary

2.1 A Motivating Example of Spatial Correlation

We consider the comovements of the returns of stock indices in developed markets in the European region. To minimize the effect of exchange rate risk, we restrict the study in the Eurozone, which consists of the countries that adopt Euro as their currency. In total, there are 11 countries with developed stock markets in the Eurozone. The data consists of the monthly simple returns of stock indices of these countries; see Table 1. Like Fama and French (2012), all returns are converted and denominated in US dollar. Since Greece adopted the Euro in the year 2000, the time period of the data spans from January 2001 to October 2013. Since all returns are denominated in US dollar, the simple return of one-month US treasury bill is used as the risk-free return.

Country	Austria	Belgium	Finland	France	Germany	Greece	Ireland	Italy	Netherlands	Portugal	Spain
Stock Index	ATX	BEL20	HEX	CAC	DAX	ASE	ISEQ	FTSEMIB	AEX	BVLX	IBEX

Table 1: The stock indices of the 11 Eurozone countries with developed stock markets.

We apply the following APT model to the monthly excess returns of the 11 stock indices:

$$r_{it} - r_{ft} = \alpha_i + \sum_{k=1}^{4} \beta_{ik} f_{kt} + \epsilon_{it}, i = 1, \ldots, 11; t = 1, \ldots, T; \tag{1}$$

where r_{it} is the return of the ith stock index in the tth month; r_{ft} is the risk-free return in the tth month; $f_{kt}, k = 1, 2, 3, 4$, are respectively the market, size, value, and momentum factors in the tth month. The four factors are defined in Fama and French (2012), and the data for the four factors are downloaded from the website of

Kenneth R. French.[1] β_{ik} is the factor loading of the ith stock index excess return on the kth factor. ϵ_{it} is the residual. To investigate potential spatial correlation among the 11 return residuals, we consider the following model for $\tilde{\epsilon}_t = (\epsilon_{1t}, \epsilon_{2t}, \ldots, \epsilon_{11t})'$:

$$\tilde{\epsilon}_t = \kappa W \tilde{\epsilon}_t + a + \tilde{\xi}_t, t = 1, 2, \ldots, T, \tag{2}$$

where $W = (w_{ij})$ is a 11×11 matrix defined as $w_{ij} := (s_i d_{ij})^{-1}$ for $i \neq j$ and $w_{ii} = 0$, where d_{ij} is the driving distance between the capital of country i and that of country j and $s_i := \sum_j d_{ij}^{-1}$; κ is a scalar parameter; a is a vector of free parameters; $\tilde{\xi}_t$ is assumed to have a normal distribution $N(0, \sigma^2 I_{11})$ with σ being an unknown parameter and I_{11} being the 11×11 identity matrix. When κ is not zero, each component of $\tilde{\epsilon}_t$ is influenced by other components to a degree dependent on their spatial distances.[2]

We fit the model (2) to the residuals and found that the spatial parameter κ for the residuals is statistically positive with a 95% confidence interval of $[0.02, 0.21]$.[3] This provides evidence that there is statistically significant spatial correlation among the residuals that is not adequately captured by the four factors in the APT model.

In addition, we carry out the following hypothesis test of the no asymptotic arbitrage constraint (i.e., zero-intercept constraint) of the APT model:

$$H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_{11} = 0;\ H_1 : \text{else}. \tag{3}$$

We find that

$$\text{The p-value of the test (3)} = 0, \tag{4}$$

which provides further evidence that the four factors may not capture the comovements of the indices returns well enough. The inadequacy of the APT model (1) for explaining the comovements of the indices returns is summarized in Table 2.

[1]These factors are called the "Fama/French European Factors" which can be downloaded at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/ftp/Europe_Factors.zip. These factors are constructed from the stocks in the developed European countries including Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, and the United Kingdom. All these factors are based on U.S. dollar denominated stock returns. The detailed description of these factors can be found at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/f-f_developed.html.

[2]Using the term $\kappa W \tilde{\epsilon}_t$ to incorporate spatial interaction is proposed in Whittle (1954) and has been widely used in spatial statistics and spatial econometrics (see, e.g., Ord (1975), Cressie (1993), and Lesage and Pace (2009)).

[3]The estimate and confidence interval for κ can be obtained by letting $K = 0$ in Eq. (35)-(38) and Eq. (122)-(123).

Questions	Answers from the data
Do the residuals in (1) have spatial correlation?	Yes, because the 95% confidence interval of κ in (2) is $[0.02, 0.21]$.
Is the no asymptotic arbitrage (zero-intercept) test (3) rejected by the data?	Yes, because of (4).

Table 2: The inadequacy of the APT model (1) for explaining the comovements of the 11 Eurozone national stock indices excess returns.

It is probable that the aforementioned unsatisfactory performance of the APT model is due to misspecification of factors. To explore this possibility, we run the Ramsey Regression Equation Specification Error Test (RESET), one of the most popular specification test for linear regression models. In our application, RESET tests whether non-linear combinations of current factors have any power in explaining the excess returns on the left-hand side of the APT regression. The intuition behind the test is that if non-linear combinations of factors have any power in explaining cross-sectional excess returns, then the factors of the APT model is misspecified. For a detailed and technical discussion of the RESET, see Ramsey (1969).

The RESET finds weak evidence of factor misspecification. Indeed, RESET indicates that among the eleven excess returns of the national stock indices, only three may benefit from additional factors. Moreover, it is not clear whether the additional factor(s) can help explain the spatial correlation observed in the residuals. Our S-APT model, to be presented in the rest of the paper, provides a unified way to address the spatial correlation in residuals. Furthermore, the no asymptotic arbitrage (i.e., zero-intercept) constraint is not rejected by the data under our S-APT model.

2.2 A Model of Spatial Interaction

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

$$r_i = \rho \sum_{j=1}^{n} w_{ij} r_j + \alpha_i + \epsilon_i, \ i = 1, \ldots, n, \tag{5}$$

where r_i is the uncertain return of asset i, α_i is a constant, and ϵ_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 ρ. 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

$$\tilde{r} = \rho W \tilde{r} + \alpha + \tilde{\epsilon}, \quad E[\tilde{\epsilon}] = 0, \quad E[\tilde{\epsilon}\tilde{\epsilon}'] = V. \tag{6}$$

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 ith row of W can simply be set to zero.

Henceforth, we assume that ρ^{-1} is not an eigenvalue of W. Then, $I_n - \rho W$ is invertible[4] and (6) can be rewritten as

$$\tilde{r} = (I_n - \rho W)^{-1}\alpha + (I_n - \rho W)^{-1}\tilde{\epsilon}, \tag{7}$$

where I_n is the $n \times n$ identity matrix. The mean and covariance matrix of $\tilde{r}$ are thus given by

$$\mu = E[\tilde{r}] = (I_n - \rho W)^{-1}\alpha, \;\; \Sigma = Cov(\tilde{r}) = (I_n - \rho W)^{-1} V (I_n - \rho W')^{-1}. \tag{8}$$

3 The Spatial Capital Asset Pricing Model

In this section we develop a spatial capital asset pricing model (S-CAPM) that generalizes 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 instruments 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 futures contract cannot be defined in the same way as that of an ordinary asset

[4] 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 ρ^{-1} is not an eigenvalue of W.

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

$$r_{n_1+i} := \frac{F_{i,1} - F_{i,0}}{F_{i,0}} \tag{9}$$

as the "nominal return" of the ith futures contract, where $F_{i,0}$ and $F_{i,1}$ are the futures prices of the ith futures contract at time 0 and time 1 (the beginning and end of the trading period), respectively, and $i = 1, \ldots, n_2$. Let $n = n_1 + n_2$ and assume that the n returns $\tilde{r} = (r_1, \ldots, r_{n_1}, r_{n_1+1}, \ldots, r_n)'$ satisfy the model (6). Then, the mean μ and covariance matrix Σ of $\tilde{r}$ are given by (8).

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. Because both μ and Σ are functions of ρ and W, the optimal portfolio weights obtained by the mean-variance analysis and the efficient frontiers are affected by spatial interaction. For example, Figure 1 shows the efficient frontiers for different values of ρ with all the other parameters in the model (6) fixed for a portfolio of ten assets. It is clear that the efficient frontiers are significantly affected by ρ.

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 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 (6), of which the first n_1 are returns of ordinary assets and the others are returns of futures contracts. Suppose $n_1 > 0$.[5] 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 only the market portfolio and the risk-free asset. Furthermore,*

(i) for the ordinary assets,

$$E[r_i] - r = \frac{Cov(r_i, r_M)}{Var(r_M)}(E[r_M] - r) = \frac{\phi_M' \Sigma \eta_i}{\phi_M' \Sigma \phi_M}(E[r_M] - r), \ i = 1, \ldots, n_1; \tag{10}$$

[5]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.

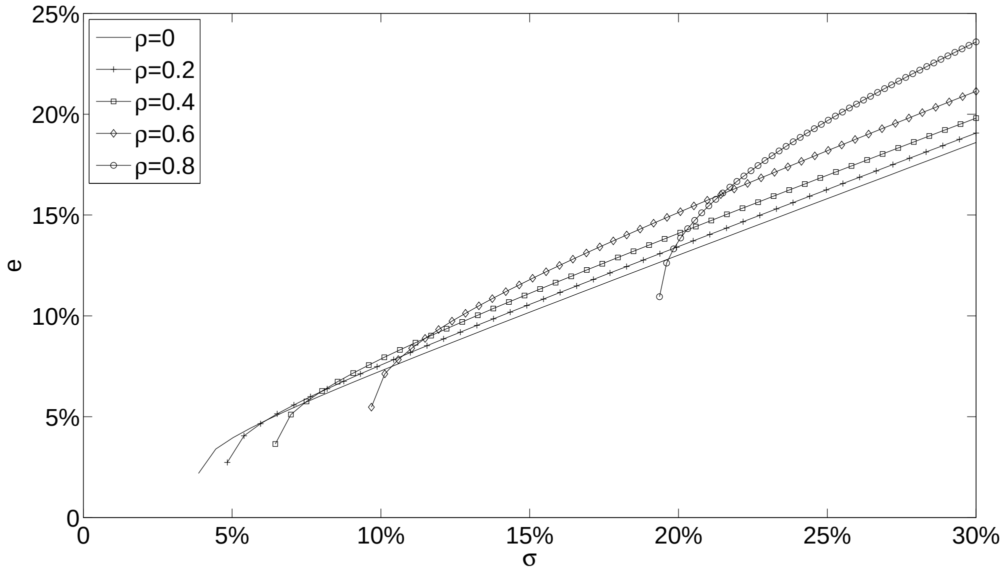


Figure 1: Efficient frontiers for $\rho = 0, 0.2, 0.4, 0.6$, and 0.8, respectively, when there is no risk-free asset. W, α, and V are specified in Appendix A. The efficient frontiers are significantly affected by ρ.

(ii) for the futures contracts,

$$E[F_{i,1}] - F_{i,0} = \frac{Cov(F_{i,1}, r_M)}{Var(r_M)}(E[r_M] - r)$$
$$= F_{i,0}\frac{\phi_M' \Sigma \eta_{n_1+i}}{\phi_M' \Sigma \phi_M}(E[r_M] - r), \ \ i = 1, \ldots, n_2, \tag{11}$$

where Σ is the covariance matrix of $\tilde{r}$; ϕ_M is the portfolio weights of the market portfolio; and η_i is the n-dimensional vector with the ith element being 1 and all other elements being 0. Define

$$1_{n_1,n_2} := (\underbrace{1, \ldots, 1}_{n_1}, \underbrace{0, \ldots, 0}_{n_2})', \tag{12}$$

then $\tilde{r} - r1_{n_1,n_2}$ is the excess asset return[6] *and the S-CAPM equations (10) and (11) are equivalent to a single equation*

$$E[\tilde{r}] - r1_{n_1,n_2} = \frac{Cov(\tilde{r}, r_M)}{Var(r_M)}(E[r_M] - r). \tag{13}$$

[6] $\tilde{r} - r1_{n_1,n_2}$ is the excess returns of the n assets in the sense that the first n_1 elements of $\tilde{r} - r1_{n_1,n_2}$ are the excess returns of the n_1 ordinary assets, and the last n_2 elements of $\tilde{r} - r1_{n_1,n_2}$ 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.

Proof. See Appendix B.1. □

By incorporating spatial interaction, the S-CAPM generalizes not only the CAPM for ordinary assets but also the CAPM for futures presented in Black (1976) and Duffie (1989, Chapter 4). The S-CAPM can also be extended to the case in which there is no risk-free asset; see Appendix B.2.

It follows from the S-CAPM equations (10) and (11) that the degree of spatial interaction represented by the parameter ρ affects asset risk premiums in equilibrium because Σ is a function of W and ρ (see (8)).

The S-CAPM implies a zero-intercept constraint on the spatial econometric models for asset returns. Consider the following spatial econometric model, in which the excess returns $\tilde{r} - r1_{n_1,n_2}$ are regressed with a spatial interaction term on the excess return of the market portfolio $r_M - r$:

$$\begin{aligned} \tilde{r} - r1_{n_1,n_2} &= \rho W(\tilde{r} - r1_{n_1,n_2}) + \bar{\alpha} + \beta(r_M - r) + \tilde{\epsilon}, \\ E[\tilde{\epsilon}] &= 0,\ Cov(r_M, \tilde{\epsilon}) = 0. \end{aligned} \tag{14}$$

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

$$\bar{\alpha} = 0. \tag{15}$$

To see this, rewrite (14) as $(I_n - \rho W)(\tilde{r} - r1_{n_1,n_2}) = \bar{\alpha} + \beta(r_M - r) + \tilde{\epsilon}$. 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 $\bar{\alpha} = (I_n - \rho W)E[(\tilde{r} - r1_{n_1,n_2}) - \frac{Cov(\tilde{r}, r_M)}{Var(r_M)}(r_M - r)]$. If the S-CAPM holds, then (13) implies $\bar{\alpha} = 0$.[7]

4 The Spatial Arbitrage Pricing Theory

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:

$$r_i = \rho \sum_{j=1}^{n} w_{ij} r_j + \alpha_i + \sum_{k=1}^{K} \beta_{ik} f_k + \epsilon_i, i = 1, \ldots, n, \tag{16}$$

[7] A spatial lag CAPM equation, which is similar to (14) with $\bar{\alpha} = 0$ and considers only ordinary assets but not futures, is *defined* in Fernandez (2011) without theoretical justification. In contrast, the present paper rigorously proves that the S-CAPM relation (13) holds (for both ordinary assets and futures) and that $\bar{\alpha}$ must be 0 in the spatial model (14) under the assumption in Theorem 3.1.

where r_i, ρ, w_{ij}, α_i, ϵ_i have the same meaning as in (6); $f_1, \ldots, f_K$ are K risk factors with $E[f_k] = 0$; and β_{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

$$\tilde{r} = \rho W \tilde{r} + \alpha + B\tilde{f} + \tilde{\epsilon}, \quad E[\tilde{f}] = 0, \quad E[\tilde{\epsilon}] = 0, \quad E[\tilde{\epsilon}\tilde{\epsilon}'] = V, \quad E[\tilde{f}\tilde{\epsilon}'] = 0. \tag{17}$$

The model (17) reduces to the classical APT when $\rho = 0$.

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 3, in the nth economy there are n_1 ordinary assets and n_2 futures contracts, where $n = n_1 + n_2$. Suppose the futures prices of the ith futures contract are $F_{i,0}^{(n)}$ and $F_{i,1}^{(n)}$ at time 0 and time 1, respectively. As in Section 3, we define the futures returns as

$$r_{n_1+i}^{(n)} = \frac{F_{i,1}^{(n)} - F_{i,0}^{(n)}}{F_{i,0}^{(n)}}, \ i = 1, \cdots, n_2. \tag{18}$$

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

$$\begin{aligned} &\tilde{r}^{(n)} = \rho^{(n)} W^{(n)} \tilde{r}^{(n)} + \alpha^{(n)} + B^{(n)} \tilde{f} + \tilde{\epsilon}^{(n)}, \text{ where} \\ &E[\tilde{f}] = 0, E[\tilde{\epsilon}^{(n)}] = 0, E[\tilde{\epsilon}^{(n)}(\tilde{\epsilon}^{(n)})'] = V^{(n)}, E[\tilde{f}(\tilde{\epsilon}^{(n)})'] = 0. \end{aligned} \tag{19}$$

The $(n+1)$th economy includes all the n risky assets in the nth economy and one extra risky asset. In the nth 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; $h_{n_1+i}^{(n)} := O_i F_{i,0}^{(n)}$, where O_i denotes the number of ith 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 (12). Then, the payoff of the zero-cost portfolio is $(h^{(n)})'(\tilde{r}^{(n)} + 1_{n_1,n_2}) = (h^{(n)})'\tilde{r}^{(n)}$,[8] because $(h^{(n)})' 1_{n_1,n_2} = 0$.

[8] 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 $(h^{(n)})'(\tilde{r}^{(n)} - r 1_{n_1,n_2})$.

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

$$E[(h^{(m_k)})'\tilde{r}^{(m_k)}] \geq \delta, \text{ for all } k, \text{ and } \lim_{k\to\infty} Var((h^{(m_k)})'\tilde{r}^{(m_k)}) = 0.^{9} \tag{20}$$

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 and there is a risk-free return r. Suppose the risk factors $\tilde{f}$ are given by

$$\tilde{f} = \tilde{g} - E[\tilde{g}], \tag{21}$$

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 (19) can then be written as

$$\tilde{r}^{(n)} - r1_{n_1,n_2} = \rho^{(n)}W^{(n)}(\tilde{r}^{(n)} - r1_{n_1,n_2}) + \bar{\alpha}^{(n)} + B^{(n)}\tilde{g} + \tilde{\epsilon}^{(n)}, \tag{22}$$

$$\bar{\alpha}^{(n)} := \alpha^{(n)} - (I_n - \rho^{(n)}W^{(n)})1_{n_1,n_2}r - B^{(n)}E[\tilde{g}]. \tag{23}$$

Theorem 4.1. *Suppose there is a risk-free return r and the risk factors $\tilde{f}$ are given by (21) where $g_1, g_2, \ldots, g_K$ are the payoffs of certain zero-cost portfolios. Suppose*

$$E[\epsilon_i^{(n)}\epsilon_j^{(n)}] = 0, \text{ for } i \neq j; \ Var(\epsilon_i^{(n)}) \leq \bar{\sigma}^2, \text{ for all } i \text{ and } n, \tag{24}$$

where $\bar{\sigma}^2$ is a fixed positive number. If there is no asymptotic arbitrage, then

$$\bar{\alpha}^{(n)} \approx 0, \tag{25}$$

or, equivalently,

$$\alpha^{(n)} \approx (I_n - \rho^{(n)}W^{(n)})1_{n_1,n_2}r + B^{(n)}E[\tilde{g}]. \tag{26}$$

The approximation (25) 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 $\bar{\alpha}^{(n)}$ whose absolute values are greater than δ.

Proof. See Appendix C.1. □

[9] In the case when there is a risk free asset with return r, the term $(h^{(m_k)})'\tilde{r}^{(m_k)}$ should be replaced by $(h^{(m_k)})'(\tilde{r}^{(m_k)} - r1_{n_1,n_2})$.

The intuition behind the theorem is that if $\tilde{g}$ are the payoffs of zero-cost portfolios, then, by (22), one can construct zero-cost portfolios with payoffs $\bar{\alpha}^{(n)} + \tilde{\epsilon}^{(n)}$ that do not carry systematic risk. If the elements of $\tilde{\epsilon}^{(n)}$ are uncorrelated and have bounded variance, then $\bar{\alpha}^{(n)}$ 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.

4.3 The Spatial Arbitrage Pricing Theory: the General Case

Theorem 4.2. *(S-APT with Both Ordinary Assets and Futures) Suppose that in the nth 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 (19). 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:*

$$\alpha^{(n)} \approx (I_n - \rho^{(n)} W^{(n)}) 1_{n_1,n_2} \lambda_0^{(n)} + B^{(n)} \lambda^{(n)}. \tag{27}$$

The precise meaning of the approximation in (27) is that there exists a positive number A such that the weighted sum of the squared pricing errors is uniformly bounded,

$$(U^{(n)})'(V^{(n)})^{-1} U^{(n)} \leq A < \infty \text{ for all } n, \tag{28}$$

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.

Proof. See Appendix C.2. □

Comparing (26) and (27), one can see that the factor risk premiums $\lambda^{(n)}$ in the S-APT can be identified as

$$\lambda^{(n)} = E[\tilde{g}], \tag{29}$$

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 $(\bar{\beta}_{i,1}^{(n)}, \bar{\beta}_{i,2}^{(n)}, \ldots, \bar{\beta}_{i,K}^{(n)})$ be the ith row of $(I_n - \rho^{(n)} W^{(n)})^{-1} B^{(n)}$. Then, (27) implies that the ordinary assets are approximately priced by

$$E[r_i^{(n)}] - \lambda_0^{(n)} \approx \sum_{k=1}^{K} \bar{\beta}_{i,k}^{(n)} \lambda_k^{(n)}, i = 1, \ldots, n_1 \tag{30}$$

and that the futures contracts are approximately priced by

$$\frac{E[F_{i,1}^{(n)}] - F_{i,0}^{(n)}}{F_{i,0}^{(n)}} \approx \sum_{k=1}^{K} \bar{\beta}_{n_1+i,k}^{(n)} \lambda_k^{(n)}, \; i = 1, \ldots, n_2. \tag{31}$$

(30) and (31) show that the expected returns of both ordinary assets and futures contracts are affected by the spatial interaction parameter ρ because, for all j and k, $\bar{\beta}_{j,k}^{(n)}$ depends on the spatial interaction terms $\rho^{(n)}$ and $W^{(n)}$.

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

$$y_i = \rho \sum_{j=1}^{n} w_{ij} y_j + \beta_0 + \sum_{k=1}^{K} \beta_k x_{ik} + \epsilon_i, i = 1, \ldots, n, \tag{32}$$

where ρ, w_{ij}, and ϵ_i have the same meaning as in (5); x_{ik} are explanatory variables; β_0 is the intercept; and $\beta_1, \ldots, \beta_K$ are coefficients in front of explanatory variables.

Although the first term in (32) of the SAR model is the same as the first term of the S-APT model (16), there are substantial differences between the two: (i) In terms of model specification, the S-APT imposes a linear constraint on model parameters ((26) or (27)), while the parameters in the SAR model are free parameters. (ii) In the SAR model (32), factors x_{ik} may be different for different i and the intercept β_0 is the same for different i; in contrast, in the S-APT model (16), factors f_k are the same for different i but the intercepts α_i depend on i.

5 Statistical Inference for S-APT

Let $\tilde{r}_t = (r_{1t}, r_{2t}, \ldots, r_{nt})'$ 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 tth period. Let r_{ft} be the risk free return in the tth period. Let $\tilde{y}_t = (y_{1t}, y_{2t}, \ldots, y_{nt})' := \tilde{r}_t - r_{ft} 1_{n_1,n_2}$ 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 tth 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 (22):

$$\begin{aligned}&\tilde{y}_t = \rho W \tilde{y}_t + \bar{\alpha} + B\tilde{g}_t + \tilde{\epsilon}_t, t = 1, 2, \ldots, T,\\&(\tilde{y}_t, \tilde{g}_t), t = 1, 2, \ldots, T, \text{ are i.i.d.,}\\&\tilde{\epsilon}_t \mid \tilde{g}_t \sim N(0, \sigma^2 I_n).\end{aligned} \tag{33}$$

The model (33) 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 these features. 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.

For the brevity of notation, we define

$$\underline{b} := (\bar{\alpha}_1, \beta_{11}, \beta_{12}, \ldots, \beta_{1K}, \ldots, \bar{\alpha}_n, \beta_{n1}, \beta_{n2}, \ldots, \beta_{nK})', \tag{34}$$

where β_{ik} is the (i, k) element of B. Denote the parameter vector of the model as $\theta := (\rho, \underline{b}', \sigma^2)'$. Let $\theta_0 = (\rho_0, \underline{b}_0', \sigma_0^2)'$ be the true model parameters.

5.1 Identifiability of Model Parameters

The model parameters θ_0 are identifiable if the spatial weight matrix W is regular (i.e., satisfying simple regularity conditions that are easy to check); see Appendix D.1 for detailed discussion. It can be easily checked that in all empirical examples of this paper, W is regular. In the rest of the section, we assume that W is regular and hence θ_0 is identifiable.[10]

[10] In fact, if W is not regular, then the elements of W satisfy $n(n+1)/2$ constraints given by (75) and (76) in Appendix D.1; 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.

5.2 Model Parameter Estimation

The model parameters can be estimated by maximum likelihood estimates (MLE). Let

$$X_t := \begin{pmatrix} 1, g_{1t}, \cdots, g_{Kt} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & 1, g_{1t}, \cdots, g_{Kt} \end{pmatrix} \in \mathbb{R}^{n \times n(K+1)}. \tag{35}$$

Then, the log likelihood function of the model is given by

$$\ell(\theta) = \ell(\rho, \underline{b}, \sigma^2) := \sum_{t=1}^{T} l(\tilde{y}_t \mid \tilde{g}_t, \theta), \text{ where} \tag{36}$$

$$\begin{aligned} l(\tilde{y}_t \mid \tilde{g}_t, \theta) = & -\frac{n}{2}\log(2\pi\sigma^2) + \frac{1}{2}\log(\det((I_n - \rho W')(I_n - \rho W))) \\ & - \frac{1}{2\sigma^2}(\tilde{y}_t - \rho W \tilde{y}_t - X_t \underline{b})'(\tilde{y}_t - \rho W \tilde{y}_t - X_t \underline{b}). \end{aligned} \tag{37}$$

Let $[\zeta, \gamma]$ be an interval such that $\zeta < 0 < \gamma$ and $I_n - \rho W$ is invertible for $\rho \in [\zeta, \gamma]$.[11] It can be shown[12] that the MLE $\hat{\theta} = (\hat{\rho}, \hat{\underline{b}}', \widehat{\sigma^2})'$ is given by

$$\hat{\rho} = \arg\max_{\rho \in [\zeta, \gamma]} \ell_c(\rho), \hat{\underline{b}} = b(\hat{\rho}), \widehat{\sigma^2} = s(\hat{\rho}),$$

where

$$\begin{aligned} \ell_c(\rho) &:= \ell(\rho, b(\rho), s(\rho)) \\ &= -\frac{nT}{2}\log(2\pi s(\rho)) + \frac{T}{2}\log(\det((I_n - \rho W')(I_n - \rho W))) - \frac{nT}{2}, \end{aligned} \tag{38}$$

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

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

For the asymptotic and small sample properties of the MLE, see Appendix D.2.

[11] For any W, because $\lim_{\rho \to 0} \det(I_n - \rho W) = 1$, there always exists an interval $[\zeta, \gamma]$ such that $\zeta < 0 < \gamma$ 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 ρ^{-1} is not an eigenvalue of W (see footnote 4). Hence, the specification of $[\zeta, \gamma]$ depends on W: (i) If W has at least two different real eigenvalues and $\omega_{\min} < 0 < \omega_{\max}$ are the minimum and maximum real eigenvalues, then $[\zeta, \gamma]$ can be chosen as an interval that lies inside $(\omega_{\min}^{-1}, \omega_{\max}^{-1})$. In particular, if the rows of W are normalized to sum up to 1, which is commonly seen in spatial econometrics literature, then $\omega_{\max} = 1$. (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.

[12] Note that for any given ρ, the original model can be rewritten as $\tilde{y}_t - \rho W \tilde{y}_t = X_t \underline{b} + \tilde{\epsilon}_t, \ \ t = 1, 2, \ldots, T$, from which the classical theory of linear regression shows that $\underline{b} = b(\rho)$ and $\sigma^2 = s(\rho)$ maximize the log likelihood function (36). Because $\ell_c(\rho) = \ell(\rho, b(\rho), s(\rho))$ and $\hat{\rho}$ maximizes $\ell_c(\rho)$, it follows that $\hat{\rho}$, $\hat{\underline{b}} = b(\hat{\rho})$, and $\widehat{\sigma^2} = s(\hat{\rho})$ maximize $\ell(\rho, \underline{b}, \sigma^2)$, i.e., they are the MLE.

5.3 Hypothesis Test and Goodness of Fit of the Model (33)

For simplicity, we assume that the factors $\tilde{g}$ are the payoffs of zero-cost tradable portfolios. In this case, Theorem 4.1 shows that the S-APT imposes an approximate zero-intercept constraint $\bar{\alpha}^{(n)} \approx 0$ (see Eq. (25)). As in the classical factor pricing literature, we test the S-APT by testing the exact zero-intercept constraint

$$H_0 : \bar{\alpha}_0 = 0; \ H_1 : \bar{\alpha}_0 \neq 0, \tag{39}$$

where $\bar{\alpha}_0$ is the true parameter in the model (33).

We can test the hypothesis using likelihood ratio test statistics. Under the null hypothesis, the likelihood ratio test statistic

$$LR = 2\left[\sum_{t=1}^{T} l(\tilde{y}_t \mid \tilde{g}_t, \hat{\theta}) - \sum_{t=1}^{T} l(\tilde{y}_t \mid \tilde{g}_t, \theta^*)\right] \tag{40}$$

has an asymptotic $\chi^2(n)$ distribution.[13] 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 θ^* estimated under the constraint that the null holds (i.e., $\bar{\alpha}_0 = 0$). The likelihood ratio test is asymptotically equivalent to the traditional tests in asset pricing, such as the Gibbons-Ross-Shanken test; see Gibbons, Ross, and Shanken (1989) as well as Chapter 1 and 2 in Hayashi (2000).

The goodness of fit of the model can be evaluated by adjusted R^2. The theoretical adjusted R^2 of the ith asset in the model (33) is defined as

$$R_i^2 = 1 - \frac{T-1}{T-K-1}\frac{Var(\epsilon_i)}{Var(y_i)}, \ i = 1, 2, \ldots, n, \tag{41}$$

where $Var(\epsilon_i) = \sigma_0^2$ and $Var(y_i)$ is equal to the ith diagonal element of the covariance matrix $(I_n - \rho_0 W)^{-1} B_0 \cdot Cov(\tilde{g}) \cdot B_0'(I_n - \rho_0 W')^{-1} + \sigma_0^2 (I_n - \rho_0 W)^{-1}(I_n - \rho_0 W')^{-1}$. The sample adjusted R^2 of the ith asset is calculated using (41) with $Var(\epsilon_i)$ and $Var(y_i)$ replaced by their respective sample counterparts.

For a simulation study of the likelihood ratio test and the adjusted R^2, see Appendix D.3.

[13] This can be shown by verifying the conditions of Proposition 7.11 in Hayashi (2000, p. 494). Let $a(\theta) := \bar{\alpha}$. Then, the Jacobian $\frac{\partial a(\theta_0)}{\partial \theta'}$ is of full row rank. We then need to verify the conditions of Proposition 7.9 in Hayashi (2000, p. 475), but it is done in the proof of Theorem D.1 in Appendix D.2 of this paper.

6 Application 1: Eurozone Stock Indices

In this section, we continue to study factor models for European stock indices returns considered in the motivating example in Section 2.1.

6.1 The Data

The data of the European stock indices returns are the same as those used in Section 2.1. We take four factors used in Fama and French (2012), namely the market factor (MKT), the size factor (SMB), the value factor (HML) and the momentum factor (MOM). The factors aforementioned have long been documented in the literature to account for the majority of comovements of equities returns. By including these four factors in the S-APT model, the empirical study is designed to more clearly reveal the contribution of spatial interaction in relation to what has been understood in the literature. Since the returns in Fama and French (2012) are sorted by corporate characteristics, while here stocks are implicitly sorted by locations/nations, one may expect some additional factors. We construct a new risk factor that is related to sovereign credit risk. Table 9 in Appendix E shows the S&P credit ratings of the 11 countries during 2001–2013, where it can be seen that Germany is the only country that has maintained top-notch AAA rating, while Greece, Ireland, Italy, Portugal, and Spain are the only countries that have breached A rating. Thus, we introduce the credit factor as the difference between returns of stock indices of countries with good credit and those with relatively bad credit:

$$g_{credit} := r_{Germany} - \frac{1}{5}(r_{Greece} + r_{Ireland} + r_{Italy} + r_{Portugal} + r_{Spain}). \tag{42}$$

Similar to the value factor constructed in Fama and French (1993), which is a proxy for the distress factor in the corporate domain, the risk factor (42) may be considered as a proxy of the sovereign version of distress risk.

6.2 Empirical Performance of S-APT and APT Models

We investigate the performance of the following six APT models and their S-APT counterparts:

- Model 1: APT with MKV, SMB, HML, and MoM factor (Fama and French (2012))

- Model 2: APT with MKV, SMB, HML, MoM, and Credit factor
- Model 3: APT with MKV, MoM, and Credit factor
- Model 4: APT with MKV, SMB, MoM, and Credit factor
- Model 5: APT with MKV, HML, MoM, and Credit factor
- Model 6: APT with MKV, SMB, HML, and Credit factor
- Model 1s: S-APT with MKV, SMB, HML, and MoM factor
- Model 2s: S-APT with MKV, SMB, HML, MoM, and Credit factor
- Model 3s: S-APT with MKV, MoM, and Credit factor
- Model 4s: S-APT with MKV, SMB, MoM, and Credit factor
- Model 5s: S-APT with MKV, HML, MoM, and Credit factor
- Model 6s: S-APT with MKV, SMB, HML, and Credit factor

An "APT" model means the type of factor model considered in Fama and French (2012) in which heterogeneous variances of residuals for different returns are assumed. The model 1 is the same as the model specified in the motivating example in Section 2.1. The S-APT model is specified in (33) with homogeneous variances for residuals. In the S-APT models, the spatial weight matrix W is defined as $W_{ij} := (s_i d_{ij})^{-1}$ for $i \neq j$ and $W_{ii} = 0$, where d_{ij} is the driving distance between the capital of country i and that of country j and $s_i := \sum_j d_{ij}^{-1}$.

The no asymptotic arbitrage (i.e., zero-intercept) hypothesis test for the APT models is specified in (3), and that for the S-APT models is specified in (39). Table 3 shows the p-value for testing the zero-intercept hypothesis, the number of parameters, the Akaike information criterion (AIC), the 95% confidence interval (C.I.) of ρ_0 (only for the S-APT models), and the 95% C.I. of κ defined in Eq. (2) for the residuals for all the above models. The domain of ρ_0 and κ in the MLE estimation is $[-2.5342, 1]$. The only models that are not rejected in the test of zero-intercept hypothesis are model 3s and 4s, which incorporate spatial interaction. The models 3s and 4s also have better performance than the APT models in terms of AIC. In general, the S-APT models seem to eliminate the spatial correlation among regression residuals more

Model	1	2	3	4	5	6
p-value	0.0	0.0	0.0002	0.0005	0.0	0.0
AIC	10172	9934	10070	9977	9961	9911
number of parameters	66	77	55	66	66	66
95% C.I. of κ for residuals	[0.02, 0.21]	[-0.01, 0.18]	[-0.01, 0.18]	[0.02, 0.20]	[-0.07, 0.13]	[-0.04, 0.15]
Model	1s	2s	3s	4s	5s	6s
p-value	0.0002	0.0099	**0.1854**	**0.2912**	0.0078	0.0002
AIC	8826	8625	**8793**	**8730**	8696	8668
number of parameters	57	68	**46**	**57**	57	57
95% C.I. of ρ_0	[0.065, 0.25]	[-0.012, 0.18]	[**0.0065**, **0.19**]	[**0.026**, **0.21**]	[-0.035, 0.16]	[0.0053, 0.19]
95% C.I. of κ for residuals	[-0.08, 0.12]	[-0.10, 0.10]	[**-0.10**, **0.11**]	[**-0.10**, **0.11**]	[-0.10, 0.10]	[-0.09, 0.11]

Table 3: The p-value for testing the no asymptotic arbitrage (i.e., zero-intercept) hypothesis, the Akaike information criterion (AIC), the number of parameters, the 95% confidence interval (C.I.) of ρ_0 (only for the S-APT models), and the 95% C.I. of κ defined in (2) for the residuals of different models for the Eurozone stock indices returns. Model 3s and Model 4s appear to perform better than the other models.

effectively than the APT models, as indicated by κ. Furthermore, for all the S-APT models that are not rejected (Models 3s and 4s), ρ_0 are found to be significantly positive. Note that adding the spatial term seems to improve the p-value and the AIC of its counterpart model without the spatial term. While the Credit, MoM, and SMB factors all seem to play a role in explaining return comovements, adding the HML factor in the model does not seem to improve model performance in terms of p-value and AIC. At last, the adjusted R^2 of fitting Model 4s with the zero-intercept constraint is reported in Figure 2.

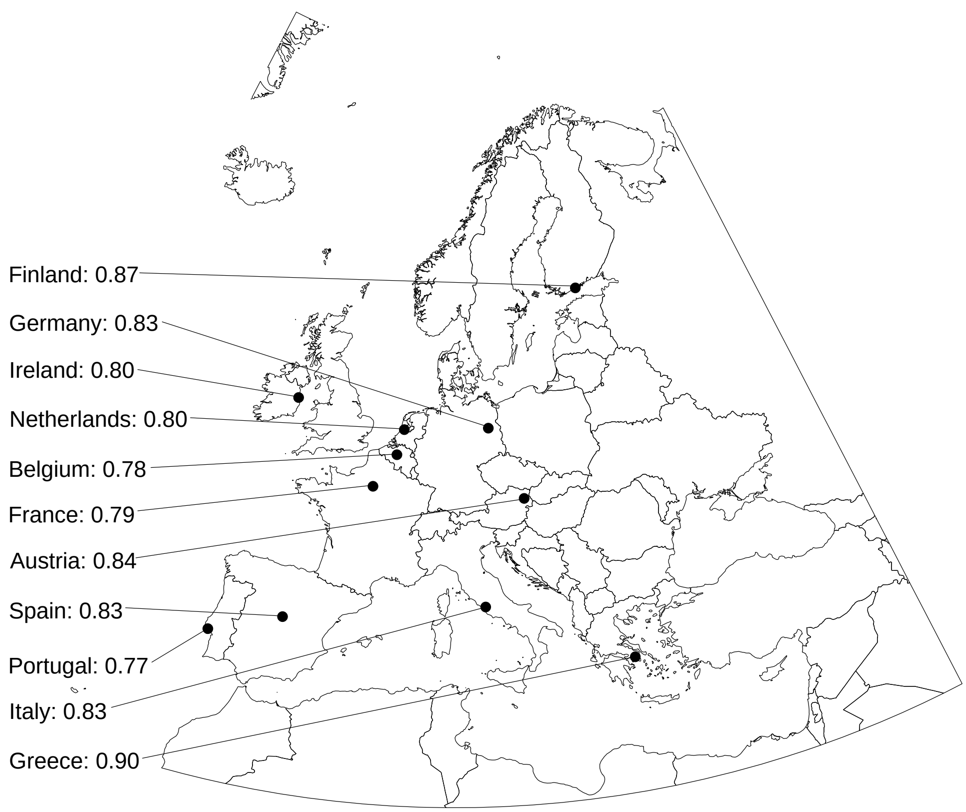


Figure 2: The adjusted R^2 of fitting the Model 4s with zero-intercept constraint to the Eurozone stock indices.

6.3 Robustness Check

The empirical results reported above seem to be robust with respect to different specifications of spatial matrix W. In Table 4 we compare the estimation and testing results using two definitions of W in the S-APT model 4s: (i) $W_{ij} := (s_i d_{ij})^{-1}$ where d_{ij} is the geographic distance[14] between the capital of the ith country and that of the jth country and $s_i := \sum_{j \neq i} d_{ij}^{-1}$. In this case, the domain of ρ_0 in MLE estimation is $[-2.6399, 1]$. (ii) $W_{ij} := (s_i d_{ij})^{-1}$ and d_{ij} is the driving distance. The numerical values of W can be found in Appendix E.

W	p-value for testing $\bar{\alpha}_0 = 0$	C.I. of ρ_0	adjusted R^2				
			Austria	Belgium	Finland	France	Germany
Geographic distance	0.2866	[0.022, 0.20]	0.8404	0.7756	0.8676	0.7858	0.8304
Driving distance	0.2912	[0.026, 0.21]	0.8405	0.7757	0.8677	0.7859	0.8305
W	AIC	adjusted R^2					
		Greece	Ireland	Italy	Netherlands	Portugal	Spain
Geographic distance	8731	0.9006	0.7950	0.8345	0.8041	0.7698	0.8296
Driving distance	8730	0.9006	0.7951	0.8346	0.8042	0.7699	0.8296

Table 4: Robustness check: the estimation and testing results under different definitions of spatial weight matrix W for the S-APT Model 4s.

7 Application 2: S&P/Case-Shiller Home Price Indices Futures

7.1 Data

The S&P/Case-Shiller Home Price Indices (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 statistical areas (MSAs) 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

[14]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.

directly traded; however, CSI Indices futures are traded at the Chicago Mercantile Exchange. There are, in total, eleven CSI Indices futures contracts; one is written on the composite 10-City CSI Index and the other ten are on the CSI Indices of ten MSAs: 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 first nearest-to-maturity futures contract to define one-month return of futures because this contract usually has better liquidity than the others. The time period of the data is from June 2006 to February 2014.

We consider three factors for the CSI indices futures. First, we construct a factor related to credit risk, as the credit risk may be a proxy of the risk of public finance (e.g. state pension schemes and infrastructure improvements). Table 10 in Appendix E shows the S&P credit ratings of the states where the 10 MSAs are located during 2006–2013. It can be seen that the credit rating of California is significantly worse than the other states, while Florida and Nevada can be chosen as representatives of states of good credit quality as both of them have a rating as good as AA+ for at least five years during the period. Following the approach of Fama and French (1993), we construct a factor related to credit risk as the difference of futures returns of MSAs in the states with relatively bad credit and those with relatively good credit,

$$g_{credit} := r_{LosAngeles} + r_{SanDiego} + r_{SanFrancisco} - (r_{Miami} + r_{LasVegas}), \qquad (43)$$

where r_{MSA} denotes the return of CSI index futures of a particular MSA. In addition to the credit factor, we consider two other factors: (i) g_{CS10f}: the monthly return of futures on S&P/Case-Shiller composite 10-City Index, which reflects the overall national residential real estate market in the United States. (ii) $g_{CS10fTr}$: the trend factor of g_{CS10f}. $g_{CS10fTr}$ in the kth month is the difference between g_{CS10f} in the kth month and the previous 12-month average of g_{CS10f}.[15] The trend factor $g_{CS10fTr}$ is inspired by a similar creation in Duan, Sun, and Wang (2012) and can be related to the notion of momentum captured by the momentum factor in Fama and French (2012). The trend factor describes the intertemporal momentum while the momentum factor focuses more on cross-sectional differences.

[15]When $k \leq 12$, the trend factor for the kth month is defined as the difference of g_{CS10f} in the kth month and the average of g_{CS10f} in the previous $k-1$ months.

7.2 Empirical Performance of S-APT and APT Models

We estimate and test three models using the same three factors defined above: One is the S-APT model specified in (33), which assumes homogeneous variance for residuals, and the other two are APT models, which are specified with homogeneous and heterogeneous residual variances respectively. APT models with heterogeneous variances seem to be common in existing theoretical and empirical works; see Gibbons, Ross, and Shanken (1989) and Fama and French (1993, 2012), among others. In the S-APT model, the spatial weight matrix W is specified based on driving distances in the same way as in Section 6.2.

We carry out the model fitting and the no asymptotic arbitrage (i.e., zero-intercept) test for the APT and S-APT models. The zero-intercept test for the APT models is specified in (3), and that for the S-APT models is specified in (39). Table 5 shows the estimation and testing results for the three models. The domain of ρ_0 in the MLE estimation for the S-APT model is $[-2.0334, 1]$. First, in testing the zero-intercept hypothesis, the S-APT model is not rejected. Second, the S-APT model outperforms the other two models in terms of AIC; in particular, the S-APT model achieves lower AIC than the APT model with homogeneous variances. Hence, incorporating spatial interaction improves the description of the comovements of futures returns. Third, for the S-APT model, ρ_0 is found to be significantly positive. Fourth, there seems to be no spatial correlation in the residuals of the S-APT model.

Figure 3a shows the sample adjusted R^2 of the S-APT model with the zero-intercept constraint $\bar{\alpha}_0 = 0$. All the sample adjusted R^2 are positive except that of New York, which is -0.35. The negative adjusted R^2 may be due to the fact that the CSI Index of New York does not reflect the overall real estate market in that area, 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.[16] Therefore, we exclude the CSI Indices futures of New York from the analysis and test the S-APT on the remaining nine CSI Indices futures. The

[16]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.

Model	S-APT	APT (heterogeneous variance)	APT (homogeneous variances)
p-value	0.114	0.0612	0.0123
AIC	-5523	-4532	-4521
Number of parameters	42	50	41
95% C.I. of ρ_0	[0.28, 0.44]	-	-
95% C.I. of κ for residuals	[-0.10, 0.11]	[0.29, 0.44]	[0.29, 0.44]

Table 5: The p-value for testing the no asymptotic arbitrage (i.e., zero-intercept) constraint, Akaike information criterion (AIC), the number of parameters, the 95% confidence interval (C.I.) of ρ_0 (only for S-APT), and the 95% C.I. of κ defined in (2) for the residuals of the three models.

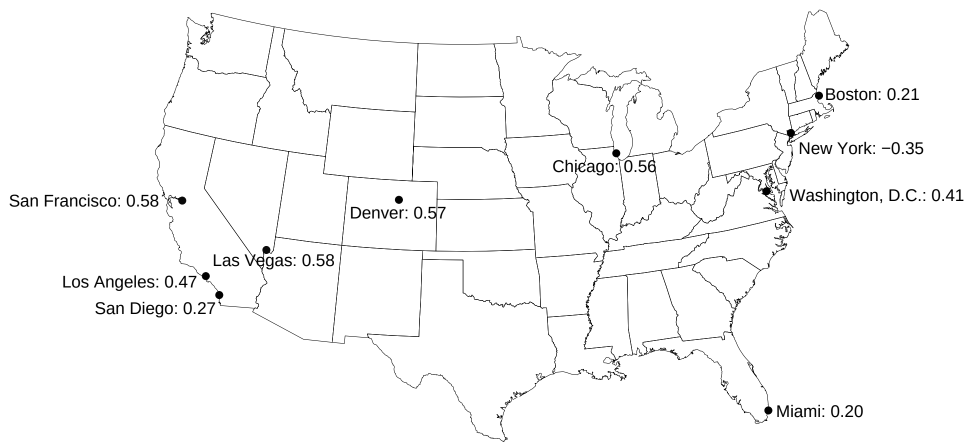
p-value of the test is 0.11, and hence the S-APT model is not rejected. The sample adjusted R^2 of fitting the remaining nine CSI Indices futures with the S-APT zero-intercept constraint is shown in Figure 3b; all the nine futures have positive adjusted R^2. ρ_0 is estimated to be 0.37 with the 95% confidence interval $[0.29, 0.45]$, which is significantly positive (the domain of ρ_0 and κ in the MLE estimation is $[-1.9897, 1]$).

7.3 Robustness Check

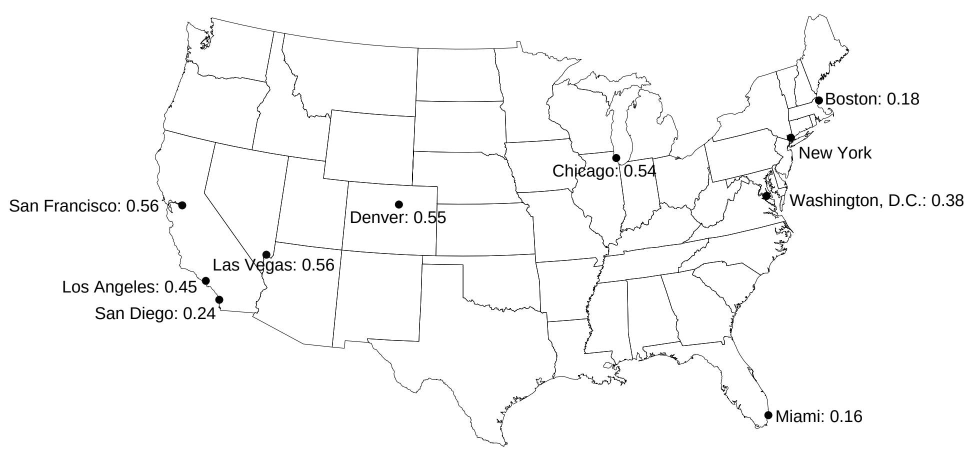
The empirical results reported above seem to be robust with respect to different specifications of spatial matrix W. Table 6 compares the model testing and estimation results for W defined by geographic distance and driving distance. The table shows that the results are robust to the specification of W. The numerical values of W can be found in Appendix E.

8 Conclusion

Although there are growing evidences that spatial interaction plays a significant role in determining prices and returns in both stock markets and real estate markets, there is as yet little work that builds explicit economic models to study the effects of spatial interaction on asset returns. In this paper, we add to the literature by studying how



(a)



(b)

Figure 3: (a) shows the adjusted R^2 of fitting the three-factor model (33) with the S-APT zero-intercept constraint $\bar{\alpha}_0 = 0$ to the 10 CSI Indices futures returns. (b) shows the adjusted R^2 of the same model fitting as in (a), except that New York is excluded from the analysis. In the model fitting, W is specified using driving distances.

(a) First robustness check: different W for ten MSAs including New York.

W	p-value for testing $\bar{\alpha}_0 = 0$	C.I. of ρ_0	AIC	adjusted R^2				
				Los Angeles	San Diego	San Francisco	Denver	Las Vegas
Geographic distance	0.1139	[0.2900, 0.4440]	-5526	0.4723	0.2772	0.5809	0.5716	0.5833
Driving distance	0.1140	[0.2835, 0.4385]	-5523	0.4705	0.2747	0.5795	0.5701	0.5819
W				adjusted R^2				
				Washington D.C.	Miami	Chicago	Boston	New York
Geographic distance				0.4133	0.1981	0.5595	0.2148	-0.3482
Driving distance				0.4113	0.1954	0.5580	0.2121	-0.3528

(b) Second robustness check: different W for nine MSAs excluding New York.

W	p-value for testing $\bar{\alpha}_0 = 0$	C.I. of ρ_0	AIC	adjusted R^2				
				Los Angeles	San Diego	San Francisco	Denver	Las Vegas
Geographic distance	0.1107	[0.3998, 0.4572]	-4934	0.4486	0.2447	0.5620	0.5523	0.5646
Driving distance	0.1100	[0.2944, 0.4516]	-4931	0.4463	0.2415	0.5602	0.5505	0.5628
W				adjusted R^2				
				Washington D.C.	Miami	Chicago	Boston	
Geographic distance				0.3869	0.1620	0.5396	0.1794	
Driving distance				0.3844	0.1586	0.5377	0.1760	

Table 6: Robustness check of the empirical results.

spatial interaction affects the risk-return relationship of financial assets. To do this, we first 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 respectively. The S-CAPM and S-APT explicitly characterize the effect of spatial interaction on the expected returns of both ordinary assets and future contracts. Next, we carry out empirical studies on the Eurozone stock indices and the futures contracts on S&P/Case-Shiller Home Price Indices using the S-APT model. Our empirical results suggest that spatial interaction is not only present but also important in explaining comovements of asset returns.

A Mean-Variance Analysis with Spatial Interaction

Assume that the n returns $\tilde{r} = (r_1, \dots, r_{n_1}, r_{n_1+1}, \dots, r_n)'$ satisfy the model (6), where the first n_1 are ordinary asset returns and the last n_2 are futures returns defined in

(9). Then, the mean μ and covariance matrix Σ of $\tilde{r}$ are given by (8).

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

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

Let r be the risk-free return. Then, the net return of the investor's portfolio is

$$\begin{aligned} r_p &= \frac{1}{w}\left(\sum_{i=1}^{n_1} u_i(1+r_i) + (w - \sum_{i=1}^{n_1} u_i)(1+r) + \sum_{j=1}^{n_2} \frac{v_j}{F_{j,0}}(F_{j,1} - F_{j,0})\right) - 1, \\ &= \frac{1}{w}\left(\sum_{i=1}^{n_1} u_i(r_i - r) + \sum_{j=1}^{n_2} v_j r_{n_1+j}\right) + r = \phi'(\tilde{r} - r1_{n_1,n_2}) + r, \end{aligned} \tag{44}$$

where $1_{n_1,n_2}$ is defined in (12). The mean and variance of r_p are given by

$$E[r_p] = h'\phi + r, \ Var(r_p) = \phi'\Sigma\phi, \text{ where } h = \mu - r1_{n_1,n_2}.$$

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

$$\min_{\phi} \ \frac{1}{2}\phi'\Sigma\phi \ \ s.t. \ \ h'\phi + r = e. \tag{45}$$

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

$$\phi^* = (e - r)\frac{\Sigma^{-1}h}{h'\Sigma^{-1}h}. \tag{46}$$

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

$$\min_{\phi} \ \frac{1}{2}\phi'\Sigma\phi \ \ s.t. \ \ \phi'\mu = e \text{ and } \phi'1_{n_1,n_2} = 1, \tag{47}$$

whose optimal solution can be shown to be

$$\phi^* = \psi + e\xi, \text{ where} \tag{48}$$

$$\psi = \frac{1}{D}(B\Sigma^{-1}1_{n_1,n_2} - A\Sigma^{-1}\mu), \xi = \frac{1}{D}(F\Sigma^{-1}\mu - A\Sigma^{-1}1_{n_1,n_2}), \tag{49}$$

$$A = \mu'\Sigma^{-1}1_{n_1,n_2}, B = \mu'\Sigma^{-1}\mu, F = 1'_{n_1,n_2}\Sigma^{-1}1_{n_1,n_2}, D = BF - A^2.$$

Because both μ and Σ are functions of ρ, the optimal portfolio weights ϕ^* and the efficient frontiers are affected by ρ. More specifically, (i) When there exists a risk-free return r, the efficient frontier is $e = r + \sigma\sqrt{H}$ and $H = (\mu - r1_{n_1,n_2})'\Sigma^{-1}(\mu - r1_{n_1,n_2}) = (\alpha - r1_{n_1,n_2})'V^{-1}(\alpha - r1_{n_1,n_2}) + 2r1'_{n_1,n_2}W'V^{-1}(\alpha - r1_{n_1,n_2})\rho + r^2 1'_{n_1,n_2}W'V^{-1}W1_{n_1,n_2}\rho^2$. Thus H is a quadratic function of ρ and the coefficient in front of ρ^2 is positive. (ii) When there is no risk-free asset, the effect of spatial interaction on efficient frontier is illustrated in Figure 1.

The model parameters used in calculating the efficient frontiers in Figure 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 $w_{ij} = \frac{1}{s_i d_{ij}}$ for $i \neq j$ and $w_{ii} = 0$, 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

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

The vector α 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 10×10 identity matrix.

B Proof of the S-CAPM Theorems

B.1 Proof of Theorem 3.1

Lemma B.1. *Let r_{mv} be any mean-variance efficient return other than the risk-free return. Then,* *(i) for any portfolio return y, it holds that*

$$E[y] - r = \frac{Cov(y, r_{mv})}{Var(r_{mv})}(E[r_{mv}] - r); \tag{50}$$

(ii) for the futures contracts, it holds that

$$E[F_{i,1}] - F_{i,0} = \frac{Cov(F_{i,1}, r_{mv})}{Var(r_{mv})}(E[r_{mv}] - r), i = 1, 2, \ldots, n_2. \tag{51}$$

Proof. Let $e_{mv} := E[r_{mv}]$. Since r_{mv} is mean-variance efficient, it follows from (46) that the portfolio weight of r_{mv} is given by

$$\phi_{mv} = (e_{mv} - r)\frac{\Sigma^{-1}h}{h'\Sigma^{-1}h}. \tag{52}$$

For any portfolio return y with the portfolio weight ϕ, it follows from (44) that

$$y = \phi'(\tilde{r} - r1_{n_1,n_2}) + r, E[y] - r = \phi'(\mu - r1_{n_1,n_2}) = \phi' h, \tag{53}$$

and then

$$Cov(y, r_{mv}) = \phi' Cov(\tilde{r})\phi_{mv} = (e_{mv} - r)\frac{\phi' h}{h'\Sigma^{-1}h}, \tag{54}$$

$$Var(r_{mv}) = Cov(r_{mv}, r_{mv}) = (e_{mv} - r)\frac{\phi'_{mv} h}{h'\Sigma^{-1}h} = \frac{(e_{mv} - r)^2}{h'\Sigma^{-1}h}. \tag{55}$$

It follows from (53), (54) and (55) that

$$\frac{Cov(y, r_{mv})}{Var(r_{mv})}(E[r_{mv}] - r) = \phi' h = E[y] - r. \tag{56}$$

In particular, consider a portfolio with all wealth w invested in the risk-free asset and a dollar-valued position u_i on the ith futures contracts. Then the portfolio return $y = u_i r_{n_1+i}/w + r$. Plugging y into (56) leads to (51). □

Proof of Theorem 3.1. Suppose that there are J investors in the economy and w_j is the initial wealth of the jth investor. Suppose that each investor selects his/her investment portfolio by solving the mean-variance problem (45) and the jth investor has a target mean return of e_j. Then, by (46), the position of the jth investor is

$$\phi^j = (e_j - r)\frac{\Sigma^{-1}h}{h'\Sigma^{-1}h}, \text{ where } h = \mu - r1_{n_1,n_2}, j = 1, 2, \ldots, J.$$

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 on futures contracts and risk free asset is zero, it follows that

$$\sum_{j=1}^{J} w_j(e_j - r)\frac{\Sigma^{-1}h}{h'\Sigma^{-1}h} = (C_1, \ldots, C_{n_1}, 0, \ldots, 0)', \tag{57}$$

$$\sum_{j=1}^{J} w_j \left(1 - (e_j - r)\frac{1'_{n_1,n_2}\Sigma^{-1}h}{h'\Sigma^{-1}h}\right) = 0, \tag{58}$$

which leads to

$$\frac{\Sigma^{-1}h}{h'\Sigma^{-1}h} = \begin{pmatrix} g \\ 0 \end{pmatrix}, \text{ where } g = \frac{1}{\sum_{j=1}^{J} w_j(e_j - r)}(C_1, \ldots, C_{n_1})'; \text{ and } \sum_{j=1}^{J} w_j = C_M. \tag{59}$$

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, (57) and (59) yield

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

which shows that the market portfolio is mean-variance efficient. Then, applying Lemma B.1 with r_{mv} being r_M leads to the conclusion of Theorem 3.1. □

B.2 The S-CAPM with Futures When There Is No Risk-free Asset

Theorem B.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 (6), of which the first n_1 are returns of ordinary assets and the others are defined in (9) 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 $Cov(r_M, r_0) = 0$, and it holds that*

(i) for the ordinary assets,

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

(ii) for the futures contracts,

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

Lemma B.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 $Cov(r_{mv}, r_0) = 0$. Furthermore,*

(i) for any portfolio return y, it holds that

$$E[y] - E[r_0] = \frac{Cov(y, r_{mv})}{Var(r_{mv})}(E[r_{mv}] - E[r_0]); \tag{60}$$

(ii) for the futures contracts, it holds that

$$E[F_{i,1}] - F_{i,0} = \frac{Cov(F_{i,1}, r_{mv})}{Var(r_{mv})}(E[r_{mv}] - E[r_0]), i = 1, 2, \ldots, n_2. \tag{61}$$

Proof. Let $e_{mv} := E[r_{mv}]$. By (48), the portfolio weight of r_{mv} is given by $\phi_{mv} = \psi + e_{mv}\xi$, where ψ and ξ is given in (49). For any portfolio return y with portfolio weight ϕ, it holds that $y = \phi'\tilde{r}$, $\phi' 1_{n_1,n_2} = 1$, and $E[y] = \phi'\mu$. Hence,

$$\begin{aligned}
&Cov(y, r_{mv}) = \phi' Cov(\tilde{r})\phi_{mv} = \phi' Cov(\tilde{r})(\psi + e_{mv}\xi)\\
=&\frac{1}{D}\left(B\phi' 1_{n_1,n_2} - A\phi'\mu + (F\phi'\mu - A\phi' 1_{n_1,n_2})e_{mv}\right)\\
=&\frac{1}{D}(B - AE[y] + Fe_{mv}E[y] - Ae_{mv}).
\end{aligned} \tag{62}$$

In particular, letting $y = r_{mv}$ in the above equation leads to $Var(r_{mv}) = \frac{1}{D}(B + Fe_{mv}^2 - 2Ae_{mv})$. Since r_{mv} is not the minimum-variance return, $e_{mv} \neq \frac{A}{F}$. Let r_0 be a mean-variance efficient return with mean $E[r_0] = \frac{B - Ae_{mv}}{A - Fe_{mv}}$. It then follows from (62) that $Cov(r_0, r_{mv}) = 0$. Since $E[r_{mv}] - E[r_0] = \frac{Fe_{mv}^2 + B - 2Ae_{mv}}{Fe_{mv} - A}$, it follows that

$$\begin{aligned}
&\frac{Cov(y, r_{mv})}{Var(r_{mv})}(E[r_{mv}] - E[r_0])\\
=&\frac{\frac{1}{D}(B - AE[y] + Fe_{mv}E[y] - Ae_{mv})}{\frac{1}{D}(B + Fe_{mv}^2 - 2Ae_{mv})} \times \frac{Fe_{mv}^2 + B - 2Ae_{mv}}{Fe_{mv} - A}\\
=&\frac{(Fe_{mv} - A)E[y] + B - Ae_{mv}}{Fe_{mv} - A} = E[y] - E[r_0],
\end{aligned} \tag{63}$$

which completes the proof of (60). Letting $y = r_1$ in (63), we have

$$E[r_1] - E[r_0] = \frac{Cov(r_1, r_{mv})}{Var(r_{mv})}(E[r_{mv}] - E[r_0]). \tag{64}$$

Letting $y = r_1 + r_{n_1+i}$, which corresponds to $\phi = (1, 0, \ldots, 0, 0, \ldots, 0, 1, \ldots, 0, 0)'$, in (63), we have

$$E[r_1] + E[r_{n_1+i}] - E[r_0] = \frac{Cov(r_1 + r_{n_1+i}, r_{mv})}{Var(r_{mv})}(E[r_{mv}] - E[r_0]). \tag{65}$$

Then, (61) follows from subtracting (64) from (65). □

Proof of Theorem B.1. When there is no risk-free asset, the mean-variance problem faced by an investor is given by (47), and the optimal portfolio weight is given by (48). 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 jth investor. Then, the position of the jth investor is

$$\phi^j = \psi + e_j\xi, \;\; j = 1, 2, \ldots, J.$$

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

$$\sum_{j=1}^{J} w_j\phi^j = \left(\sum_{j=1}^{J} w_j\right)\psi + \left(\sum_{j=1}^{J} w_j e_j\right)\xi = (C_1, \ldots, C_{n_1}, 0, \ldots, 0)'. \tag{66}$$

By (66), $\sum_{j=1}^{J} w_j 1'_{n_1,n_2}\phi^j = 1'_{n_1,n_2}(C_1, \ldots, C_{n_1}, 0, \ldots, 0)' = C_M$, which in combination with $1'_{n_1,n_2}\phi^j = 1$ leads to $\sum_{j=1}^{J} w_j = C_M$. Let $e_M = \sum_{j=1}^{J} w_j e_j / C_M$. Then, it follows from (66) that

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

which shows that the market portfolio is mean-variance efficient with a target mean return e_M. Then, the conclusion in Theorem B.1 follows by applying Lemma B.2 with the market portfolio return r_M being r_{mv}. □

C Proof of the S-APT Theorems

C.1 Proof of Theorem 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 (17) as

$$(I_n - \rho^{(n)} W^{(n)})(\tilde{r}^{(n)} - r 1_{n_1,n_2}) = \bar{\alpha}^{(n)} + B^{(n)} \tilde{F} + \tilde{\epsilon}^{(n)}.$$

Let η_j be the jth 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 $\eta_j'(I_n - \rho^{(n)} W^{(n)})(\tilde{r}^{(n)} - r 1_{n_1,n_2}) - \eta_j' B^{(n)} \tilde{F}$, 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 $\bar{\sigma}^2$; 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 δ and variance less than $\bar{\sigma}^2 / N(n, \delta)$. 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 δ and diminishing variances, constituting an asymptotic arbitrage opportunity. Therefore, if no asymptotic arbitrage opportunities exist, then there exists a number N_δ not depending on n such that $N(n, \delta) < N_\delta$ for all n. Since δ can be arbitrarily small, we conclude that $\bar{\alpha}^{(n)} \approx 0$. □

C.2 Proof of Theorem 4.2

Proof. Since $I_n - \rho^{(n)} W^{(n)}$ is invertible, the model (19) can be written as

$$\tilde{r}^{(n)} = Q^{(n)} \alpha^{(n)} + Q^{(n)} B^{(n)} \tilde{f} + Q^{(n)} \tilde{\epsilon}^{(n)}, \text{ where } Q^{(n)} = (I_n - \rho^{(n)} W^{(n)})^{-1}. \tag{67}$$

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

$$\hat{\alpha} = Q\alpha, \hat{B} = QB, \tilde{\varepsilon} = Q\tilde{\epsilon}, \Omega = QVQ'. \tag{68}$$

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

$$C^{-1}\hat{\alpha} = C^{-1} 1_{n_1,n_2} \lambda_0 + C^{-1}\hat{B}\lambda + u,$$

where u satisfies the orthogonality condition:

$$0 = \hat{B}'(C')^{-1}u, \tag{69}$$

$$0 = 1'_{n_1,n_2}(C')^{-1}u. \tag{70}$$

Define the pricing errors $v := \hat{\alpha} - 1_{n_1,n_2}\lambda_0 - \hat{B}\lambda = Cu$. Then, by (69) and (70), we have

$$0 = \hat{B}'(C')^{-1}u = \hat{B}'(C')^{-1}C^{-1}v = \hat{B}'\Omega^{-1}v, \tag{71}$$

$$0 = 1'_{n_1,n_2}(C')^{-1}u = 1'_{n_1,n_2}(C')^{-1}C^{-1}v = 1'_{n_1,n_2}\Omega^{-1}v. \tag{72}$$

Consider the portfolio $h = \Omega^{-1}v(v'\Omega^{-1}v)^{-1}$. By (72), h is a zero-cost portfolio. By (71), the payoff of the zero-cost portfolio is

$$\begin{aligned} h'\tilde{r} &= h'(Q\alpha + QB\tilde{f} + Q\tilde{\epsilon}) = h'(\hat{\alpha} + \hat{B}\tilde{f} + \tilde{\varepsilon}) = h'\hat{\alpha} + (v'\Omega^{-1}v)^{-1}v'\Omega^{-1}\hat{B}\tilde{f} + h'\tilde{\varepsilon} \\ &= h'\hat{\alpha} + h'\tilde{\varepsilon}, \end{aligned}$$

whose expectation and variance are

$$\begin{aligned} E[h'\tilde{r}] &= h'\hat{\alpha} = (v'\Omega^{-1}v)^{-1}v'\Omega^{-1}(1_{n_1,n_2}\lambda_0 + \hat{B}\lambda + v) = 1, \\ Var(h'\tilde{r}) &= h'\Omega h = (v'\Omega^{-1}v)^{-2}v'\Omega^{-1}\Omega\Omega^{-1}v = (v'\Omega^{-1}v)^{-1} \\ &= [(\hat{\alpha} - 1_{n_1,n_2}\lambda_0 - \hat{B}\lambda)'(Q')^{-1}V^{-1}(Q)^{-1}(\hat{\alpha} - 1_{n_1,n_2}\lambda_0 - \hat{B}\lambda)]^{-1} \\ &= [(\alpha - (I_n - \rho W)1_{n_1,n_2}\lambda_0 - B\lambda)'V^{-1}(\alpha - (I_n - \rho W)1_{n_1,n_2}\lambda_0 - B\lambda)]^{-1} \\ &= (U'V^{-1}U)^{-1}. \end{aligned}$$

Therefore, if (28) 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 $Q, \hat{\alpha}, \hat{B}, \tilde{\varepsilon}$, and Ω be defined in (67) and (68). Since Ω 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

$$0 = \hat{B}'(C')^{-1}u = \hat{B}'(C')^{-1}C^{-1}v = \hat{B}'\Omega^{-1}v. \tag{73}$$

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 (68) and (73), the payoff of the portfolio is

$$\begin{aligned}
&h'(1_{n_1,n_2} + \tilde{r}) - h'1_{n_1,n_2}(1 + r) = h'(\tilde{r} - r1_{n_1,n_2}) = h'(Q\alpha + QB\tilde{f} + Q\tilde{\epsilon} - r1_{n_1,n_2}) \\
&= h'(\hat{\alpha} + \hat{B}\tilde{f} + \tilde{\varepsilon} - r1_{n_1,n_2}) = h'(\hat{\alpha} - r1_{n_1,n_2}) + (v'\Omega^{-1}v)^{-1}v'\Omega^{-1}\hat{B}\tilde{f} + h'\tilde{\varepsilon} \\
&= h'(\hat{\alpha} - r1_{n_1,n_2}) + h'\tilde{\varepsilon},
\end{aligned}$$

whose mean and variance are

$$\begin{aligned}
&E[h'(\tilde{r} - r1_{n_1,n_2})] = h'(\hat{\alpha} - r1_{n_1,n_2}) = (v'\Omega^{-1}v)^{-1}v'\Omega^{-1}(\hat{B}\lambda + v) = 1, \\
&Var(h'(\tilde{r} - r1_{n_1,n_2})) = Var(h'\tilde{\varepsilon}) = h'\Omega h = (v'\Omega^{-1}v)^{-2}v'\Omega^{-1}\Omega\Omega^{-1}v \\
&= (v'\Omega^{-1}v)^{-1} = [(\hat{\alpha} - r1_{n_1,n_2} - \hat{B}\lambda)'(Q')^{-1}V^{-1}(Q)^{-1}(\hat{\alpha} - r1_{n_1,n_2} - \hat{B}\lambda)]^{-1} \\
&= [(\alpha - (I_n - \rho W)1_{n_1,n_2}r - B\lambda)'V^{-1}(\alpha - (I_n - \rho W)1_{n_1,n_2}r - B\lambda)]^{-1}.
\end{aligned}$$

Therefore, if (28) (with $\lambda_0^{(n)}$ replaced by r) is violated, then the variance of $h'(\tilde{r} - r1_{n_1,n_2})$ vanishes along some subsequence, and an asymptotic arbitrage opportunity exists. The proof is thus completed. □

D Econometric Tools for Implementing the S-APT Model (33)

D.1 Identifiability of Model Parameters

We need the following mild assumptions regarding $\tilde{g}_t$ for proving Proposition D.1 and Proposition D.2:

Assumption D.1. *We assume that $\tilde{g}_t$ satisfies the following mild technical conditions:*

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

(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.

Let $\theta_0 = (\rho_0, \underline{b}_0', \sigma_0^2)'$ be the true model parameters that lie in the interior of the parameter space Θ defined by:

$$\Theta := [\zeta, \gamma] \times [-\delta_b, \delta_b]^{n\times(K+1)} \times [\delta_s^{-1}, \delta_s], \tag{74}$$

where $\zeta < 0 < \gamma, \delta_b > 0, \delta_s > 0$ are constants and $I_n - \rho W$ is invertible for $\rho \in [\zeta, \gamma]$.[17]

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

Assumption D.2. *W is non-negative; $W \neq 0$; and the diagonal elements of W are all equal to zero.*

Let $l(\tilde{y}_t \mid \tilde{g}_t, \theta)$ be defined in (37). We recall the following definition of identifiability; see, e.g., Neway and McFadden (1994, Lemma 2.2).

Definition D.1. *θ_0 is identifiable if for any $\theta \neq \theta_0$ and $\theta \in \Theta$ it holds that $P(l(\tilde{y}_t \mid \tilde{g}_t, \theta) \neq l(\tilde{y}_t \mid \tilde{g}_t, \theta_0)) > 0$.*

It turns out that the identifiability of θ_0 depends largely on the property of the spatial weight matrix W. In particular, we need the following definition for W:

Definition D.2. *The spatial weight matrix W is regular if there exist no $c_1 > 0$ and $c_2 \geq 0$ such that*

$$\sum_{k=1}^{n} W_{ki}^2 = c_1, \ \forall i = 1, \ldots, n, \tag{75}$$

$$\sum_{k=1}^{n} W_{ki} W_{kj} = c_2(W_{ij} + W_{ji}), \ \forall 1 \leq i < j \leq n. \tag{76}$$

If W is not regular, then the pair of constants (c_1, c_2) that satisfy (75) and (76) are unique. Indeed, by Assumption D.2, there exists $i < j$ such that $W_{ij} + W_{ji} > 0$; hence, it follows from (76) that c_2 is uniquely determined by W. Apparently, c_1 is also unique. We have the following proposition regarding the identifiability of θ_0.

Proposition D.1. *Any θ_0 is identifiable if W is regular. More generally, a particular θ_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 (75) and (76),

[17]See footnote 11 for more details on the specification of the interval $[\zeta, \gamma]$.

and one of the following conditions holds:

$$\rho_0 = -\frac{c_2}{c_1}; \tag{77}$$

$$\rho_0 \neq -\frac{c_2}{c_1} \text{ and } \frac{1 - c_2\rho_0}{c_2 + c_1\rho_0} = \rho_0; \tag{78}$$

$$\rho_0 \neq -\frac{c_2}{c_1} \text{ and } \frac{1 - c_2\rho_0}{c_2 + c_1\rho_0} \neq \rho_0 \text{ and } \theta_* := (\rho_*, \bar{\alpha}_*, B_*, \sigma_*^2) \notin \Theta, \tag{79}$$

where $\rho_* := \frac{1-c_2\rho_0}{c_2+c_1\rho_0}$, $\sigma_*^2 := \sigma_0^2 \frac{c_2^2+c_1}{(c_2+c_1\rho_0)^2}$, $\bar{\alpha}_* := \frac{\sigma_*^2}{\sigma_0^2}(I_n - \rho_* W')^{-1}(I_n - \rho_0 W')\bar{\alpha}_0$, *and* $B_* := \frac{\sigma_*^2}{\sigma_0^2}(I_n - \rho_* W')^{-1}(I_n - \rho_0 W')B_0$.

Proof. θ_0 is not identifiable if and only if there exists $\theta \in \Theta$ and $\theta \neq \theta_0$ such that

$$P(l(\tilde{y}_t \mid \tilde{g}_t; \theta) = l(\tilde{y}_t \mid \tilde{g}_t; \theta_0)) = 1. \tag{80}$$

It follows from the part (ii) of Assumption D.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, (80) implies that

$$l(\tilde{y} \mid \tilde{g}; \theta) = l(\tilde{y} \mid \tilde{g}; \theta_0),\ \forall (\tilde{y}, \tilde{g}) \in \mathbb{R}^n \times \mathbb{A}.$$

By (37), we have

$$l(\tilde{y} \mid \tilde{g}; \theta) = \tilde{y}' C_1(\theta)\tilde{y} + \tilde{g}' C_2(\theta)\tilde{g} + \tilde{g}' C_3(\theta)\tilde{y} + C_4(\theta)'\tilde{y} + C_5(\theta)'\tilde{g} + C_6(\theta), \tag{81}$$

where

$$C_1(\theta) = -\frac{1}{2\sigma^2}(I_n - \rho W')(I_n - \rho W)$$

$$C_2(\theta) = -\frac{1}{2\sigma^2}B'B \tag{82}$$

$$C_3(\theta) = \frac{1}{\sigma^2}B'(I_n - \rho W) \tag{83}$$

$$C_4(\theta) = \frac{1}{\sigma^2}(I_n - \rho W')\bar{\alpha} \tag{84}$$

$$C_5(\theta) = -\frac{1}{\sigma^2}B'\bar{\alpha} \tag{85}$$

$$C_6(\theta) = -\frac{n}{2}\log(2\pi\sigma^2) + \frac{1}{2}\log(\det((I_n - \rho W')(I_n - \rho W))) - \frac{1}{2\sigma^2}\bar{\alpha}'\bar{\alpha}. \tag{86}$$

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

$$C_i(\theta) = C_i(\theta_0),\ i = 1, 2, \ldots, 6. \tag{87}$$

Since $I_n - \rho W$ is invertible, (83) and (84) imply that

$$B' = \sigma^2 C_3(\theta)(I_n - \rho W)^{-1},\ \bar{\alpha} = \sigma^2 (I_n - \rho W')^{-1} C_4(\theta). \tag{88}$$

Then, plugging (88) into (82), (85), and (86), we obtain

$$\begin{aligned}
C_2(\theta) &= \frac{1}{4} C_3(\theta) C_1(\theta)^{-1} C_3(\theta)', \\
C_5(\theta) &= \frac{1}{2} C_3(\theta) C_1(\theta)^{-1} C_4(\theta), \\
C_6(\theta) &= -\log((2\pi)^{\frac{n}{2}} \det(-2C_1(\theta))^{-\frac{1}{2}}) + \frac{1}{4} C_4(\theta)' C_1(\theta)^{-1} C_4(\theta).
\end{aligned}$$

Hence, (87) is equivalent to

$$C_i(\theta) = C_i(\theta_0),\ i = 1, 3, 4. \tag{89}$$

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 (80) holds. Then, (89) holds. $C_1(\theta) = C_1(\theta_0)$ is equivalent to that

$$\left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma^2}\right) I_n - \left(\frac{\rho_0}{\sigma_0^2} - \frac{\rho}{\sigma^2}\right)(W' + W) + \left(\frac{\rho_0^2}{\sigma_0^2} - \frac{\rho^2}{\sigma^2}\right) W'W = 0. \tag{90}$$

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

$$\frac{1}{\sigma_0^2} - \frac{1}{\sigma^2} + \left(\frac{\rho_0^2}{\sigma_0^2} - \frac{\rho^2}{\sigma^2}\right) \sum_{k=1}^{n} W_{ki}^2 = 0,\ i = 1, 2, \ldots, n. \tag{91}$$

For $i < j$, considering the (i,j)-element of the matrices on both sides of (90), we obtain

$$-\left(\frac{\rho_0}{\sigma_0^2} - \frac{\rho}{\sigma^2}\right)(W_{ij} + W_{ji}) + \left(\frac{\rho_0^2}{\sigma_0^2} - \frac{\rho^2}{\sigma^2}\right) \sum_{k=1}^{n} W_{ki} W_{kj} = 0,\ \forall 1 \leq i < j \leq n. \tag{92}$$

Suppose that $\frac{\rho_0^2}{\sigma_0^2} - \frac{\rho^2}{\sigma^2} = 0$, then (91) implies $\sigma = \sigma_0$, which, together with (92) and that there exists $i \neq j$ such that $W_{ij} + W_{ji} > 0$ (by Assumption D.2), implies that $\rho = \rho_0$. Then, since $I_n - \rho W$ is invertible, $C_3(\theta) = C_3(\theta_0)$ and (83) imply $B = B_0$; $C_4(\theta) = C_4(\theta_0)$ and (84) 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 (91) and that there exists i such that $\sum_{k=1}^n W_{ki}^2 > 0$ (by Assumption D.2), it follows that $\sigma_0 > \sigma$. Hence, (91) and (92) imply that

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

where

$$c_1 = -\frac{\frac{1}{\sigma_0^2} - \frac{1}{\sigma^2}}{\frac{\rho_0^2}{\sigma_0^2} - \frac{\rho^2}{\sigma^2}} > 0 \text{ and } c_2 = \frac{\frac{\rho_0}{\sigma_0^2} - \frac{\rho}{\sigma^2}}{\frac{\rho_0^2}{\sigma_0^2} - \frac{\rho^2}{\sigma^2}} \ge 0, \tag{93}$$

which contradicts to the assumption that W is regular.

(ii) Suppose that W is not regular and corresponds to $c_1 > 0$ and $c_2 \ge 0$ in (75) and (76). Suppose for the sake of contradiction that there exists $\theta \neq \theta_0$ such that (80) holds. Then, by the same argument in case (i) and by the uniqueness of (c_1, c_2), $\sigma \neq \sigma_0$ and (ρ, σ) must satisfy the equations in (93) and hence must be a solution to the following two equations

$$c_1(\frac{\rho_0^2}{\sigma_0^2} - \frac{\rho^2}{\sigma^2}) = -\frac{1}{\sigma_0^2} + \frac{1}{\sigma^2} \text{ and } c_2(\frac{\rho_0^2}{\sigma_0^2} - \frac{\rho^2}{\sigma^2}) = \frac{\rho_0}{\sigma_0^2} - \frac{\rho}{\sigma^2}.$$

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

$$(c_2 + c_1\rho_0)\rho^2 - (c_1\rho_0^2 + 1)\rho - (c_2\rho_0^2 - \rho_0) = 0 \text{ and } \sigma^2 = \frac{1 + c_1\rho^2}{1 + c_1\rho_0^2}\sigma_0^2. \tag{94}$$

If $c_2 + c_1\rho_0 = 0$, then the above system of equations has a unique solution (ρ_0, σ_0); otherwise, the above equations have two solutions (ρ_0, σ_0) and $(\frac{1-c_2\rho_0}{c_2+c_1\rho_0}, \sigma_0\sqrt{\frac{c_2^2+c_1}{(c_2+c_1\rho_0)^2}})$.

(ii.1) Suppose that (77) or (78) holds, then the two equations in (94) have a unique solution (ρ_0, σ_0), and hence the two equations in (93) do not have a solution $(\rho, \sigma) \neq (\rho_0, \sigma_0)$, which leads to a contradiction.

(ii.2) Suppose that (79) holds, then $(\frac{1-c_2\rho_0}{c_2+c_1\rho_0}, \sigma_0\sqrt{\frac{c_2^2+c_1}{(c_2+c_1\rho_0)^2}})$ is the unique solution to (93); hence, $\rho = \frac{1-c_2\rho_0}{c_2+c_1\rho_0} = \rho_*$ and $\sigma = \sigma_0\sqrt{\frac{c_2^2+c_1}{(c_2+c_1\rho_0)^2}} = \sigma_*$. Since $C_3(\theta) = C_3(\theta_0)$ and $C_4(\theta) = C_4(\theta_0)$, it follows that $\bar{\alpha} = \bar{\alpha}_*$ and $B = B_*$. Hence, $\theta = \theta_*$. However, $\theta \in \Theta$ but $\theta_* \notin \Theta$, 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 $c_2 \ge 0$, 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 $\theta_* \in \Theta$. Then, by the definition of θ_*, it holds that $\theta_* \neq \theta_0$ and $C_i(\theta_*) = C_i(\theta_0)$ for $i = 1, 3,$

and 4, which further implies that $C_i(\theta_*) = C_i(\theta_0)$ 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 θ_0 is identifiable. □

Proposition D.1 is equivalent to the following statement: A particular θ_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 (75) and (76), 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 $\theta_* \in \Theta$.

D.2 Asymptotic and Small Sample Properties of the MLE

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

Proposition D.2. *Define*

$$Q_0(\theta) := E[l(\tilde{y}_t \mid \tilde{g}_t, \theta)], \quad \hat{Q}_T(\theta) := \frac{1}{T}\sum_{t=1}^{T} l(\tilde{y}_t \mid \tilde{g}_t, \theta). \tag{95}$$

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

$$s(\tilde{y}_t, \tilde{g}_t; \theta) := \frac{\partial l(\tilde{y}_t \mid \tilde{g}_t, \theta)}{\partial \theta} \text{ and } H(\tilde{y}_t, \tilde{g}_t; \theta) := \frac{\partial^2 l(\tilde{y}_t \mid \tilde{g}_t, \theta)}{\partial \theta \partial \theta'}. \tag{96}$$

Then, the following statements hold:

(i) $Q_0(\theta)$ *is uniquely maximized at* θ_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* Θ.

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

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

$$\begin{bmatrix} -\operatorname{tr}(W(I_n - \rho W)^{-1} W (I_n - \rho W)^{-1}) - \frac{1}{\sigma^2}\tilde{y}_t' W' W \tilde{y}_t & -\frac{1}{\sigma^2}\tilde{y}_t' W' X_t & -\frac{1}{\sigma^4}\tilde{y}_t' W' \tilde{\xi}_t \\ -\frac{1}{\sigma^2} X_t' W \tilde{y}_t & -\frac{1}{\sigma^2} X_t' X_t & -\frac{1}{\sigma^4} X_t' \tilde{\xi}_t \\ -\frac{1}{\sigma^4}\tilde{\xi}_t' W \tilde{y}_t & -\frac{1}{\sigma^4}\tilde{\xi}_t' X_t & \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\tilde{\xi}_t'\tilde{\xi}_t \end{bmatrix}, \tag{97}$$

where $\tilde{\xi}_t := (I_n - \rho W)\tilde{y}_t - X_t \underline{b}$, *and* $\operatorname{tr}(\cdot)$ *denotes the matrix trace.*

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

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

(viii) There is a neighborhood $\mathcal{N}$ of θ_0 such that $E[\sup_{\theta\in\mathcal{N}} \|H(\tilde{y}_t, \tilde{g}_t; \theta)\|] < \infty$.

Proof. Since $\det((I_n - \rho W')(I_n - \rho W))$ is equal to a polynomial of ρ, it follows from (81) that $\hat{Q}_T(\theta)$ is twice continuously differentiable on the interior of Θ. The proof of 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 (33) and part (i) of Assumption D.1 that $E[\|\tilde{y}_t\|^2] < \infty$. Hence, (81) 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 $g(\tilde{y}_t, \tilde{g}_t) := \frac{f(\tilde{y}_t|\tilde{g}_t,\theta)}{f(\tilde{y}_t|\tilde{g}_t,\theta_0)}$. Since θ_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

$$E[l(\tilde{y}_t \mid \tilde{g}_t, \theta_0) - l(\tilde{y}_t \mid \tilde{g}_t, \theta)] = E[-\log g(\tilde{y}_t, \tilde{g}_t)] > -\log E[g(\tilde{y}_t, \tilde{g}_t)]. \tag{98}$$

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 (98) implies that $Q_0(\theta)$ has a unique maximizer θ_0.

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

$$E[\sup_{\theta\in\Theta} |l(\tilde{y}_t \mid \tilde{g}_t; \theta)|] < \infty. \tag{99}$$

By (81), $l(\tilde{y}_t \mid \tilde{g}_t; \theta) = \sum_{i,j} a_{ij}(\theta) y_{it} y_{jt} + \sum_{i,j} b_{ij}(\theta) g_{it} g_{jt} + \sum_{i,j} c_{ij}(\theta) y_{it} g_{jt} + \sum_{i=1}^{n} d_i(\theta) y_{it} + \sum_{i=1}^{K} e_i(\theta) g_{it} + C_6(\theta)$, 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 Θ 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 θ) 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 $\tilde{\xi}_t := \tilde{y}_t - \rho W \tilde{y}_t - X_t \underline{b}$. By Jacobi's formula of matrix calculus,

$$\frac{d}{d\rho} \det((I_n - \rho W')(I_n - \rho W)) = -2 \det((I_n - \rho W')(I_n - \rho W)) \operatorname{tr}((I_n - \rho W)^{-1} W),$$

where $\operatorname{tr}(\cdot)$ denotes the trace of a matrix. Hence,

$$\frac{\partial l(\tilde{y}_t \mid \tilde{g}_t, \theta)}{\partial \rho} = -\operatorname{tr}((I_n - \rho W)^{-1} W) + \frac{1}{\sigma^2} \tilde{\xi}_t' W \tilde{y}_t. \tag{100}$$

By simple algebra, we have

$$\frac{\partial l(\tilde{y}_t \mid \tilde{g}_t, \theta)}{\partial \underline{b}} = \frac{1}{\sigma^2} X_t' \tilde{\xi}_t, \ \frac{\partial l(\tilde{y}_t \mid \tilde{g}_t, \theta)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \tilde{\xi}_t' \tilde{\xi}_t. \tag{101}$$

Hence,

$$\begin{aligned} E\left[\frac{\partial l(\tilde{y}_t \mid \tilde{g}_t, \theta_0)}{\partial \rho}\right] &= -\operatorname{tr}((I_n - \rho_0 W)^{-1} W) + \frac{1}{\sigma_0^2} E[\tilde{\xi}_t' W (I_n - \rho_0 W)^{-1} (\bar{\alpha}_0 + B_0 \tilde{g}_t + \tilde{\xi}_t)] \\ &= -\operatorname{tr}((I_n - \rho_0 W)^{-1} W) + \frac{1}{\sigma_0^2} E[\tilde{\xi}_t' W (I_n - \rho_0 W)^{-1} \tilde{\xi}_t] \\ &= -\operatorname{tr}((I_n - \rho_0 W)^{-1} W) + \frac{1}{\sigma_0^2} E[\operatorname{tr}(W (I_n - \rho_0 W)^{-1} \tilde{\xi}_t \tilde{\xi}_t')] \\ &= -\operatorname{tr}((I_n - \rho_0 W)^{-1} W) + \frac{1}{\sigma_0^2} \sigma_0^2 \operatorname{tr}(W (I_n - \rho_0 W)^{-1}) = 0. \end{aligned}$$

Also, by (101),

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

which completes the proof of part (iv).

(v) For brevity of notation, we use $l(\theta)$ to denote $l(\tilde{y}_t \mid \tilde{g}_t, \theta)$ in the sequel. By (100) and (101), we have

$$\frac{\partial^2 l(\theta)}{\partial \rho^2} = -\operatorname{tr}(W (I_n - \rho W)^{-1} W (I_n - \rho W)^{-1}) - \frac{1}{\sigma^2} \tilde{y}_t' W' W \tilde{y}_t, \tag{102}$$

$$\frac{\partial^2 l(\theta)}{\partial \rho \partial \underline{b}} = -\frac{1}{\sigma^2} X_t' W \tilde{y}_t, \frac{\partial^2 l(\theta)}{\partial \rho \partial \sigma^2} = -\frac{1}{\sigma^4} \tilde{\xi}_t' W \tilde{y}_t, \frac{\partial^2 l(\theta)}{\partial \underline{b} \partial \underline{b}'} = -\frac{1}{\sigma^2} X_t' X_t, \tag{103}$$

$$\frac{\partial^2 l(\theta)}{\partial \sigma^2 \partial \underline{b}} = -\frac{1}{\sigma^4} X_t' \tilde{\xi}_t, \frac{\partial^2 l(\theta)}{\partial \sigma^2 \partial \sigma^2} = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6} \tilde{\xi}_t' \tilde{\xi}_t, \tag{104}$$

which completes the proof.

(vi) Let

$$C := W (I_n - \rho_0 W)^{-1} \tag{105}$$

and let C_{ij} be the (i, j) element of C. Then by (102), we have

$$\begin{aligned} & E[\frac{\partial^2 l(\theta_0)}{\partial \rho^2} \mid \tilde{g}_t] \\ = & -\operatorname{tr}(C^2) - \frac{1}{\sigma_0^2} E[((I_n - \rho_0 W)^{-1} (X_t \underline{b}_0 + \tilde{\epsilon}_t))' W' W (I_n - \rho_0 W)^{-1} (X_t \underline{b}_0 + \tilde{\epsilon}_t) \mid \tilde{g}_t] \\ = & -\operatorname{tr}(C^2) - \frac{1}{\sigma_0^2} \underline{b}_0' X_t' C' C X_t \underline{b}_0 - \frac{1}{\sigma_0^2} E[\tilde{\epsilon}_t' C' C \tilde{\epsilon}_t] \\ = & -\operatorname{tr}(C^2) - \frac{1}{\sigma_0^2} \underline{b}_0' X_t' C' C X_t \underline{b}_0 - \operatorname{tr}(C' C). \end{aligned} \tag{106}$$

By (100) and simple algebra,

$$E[\left(\frac{\partial l(\theta_0)}{\partial\rho}\right)^2 \mid \tilde{g}_t] = E[(-\operatorname{tr}(C) + \frac{1}{\sigma_0^2}\tilde{\epsilon}_t' W\tilde{y}_t)^2 \mid \tilde{g}_t]$$

$$= E[(-\operatorname{tr}(C) + \frac{1}{\sigma_0^2}\tilde{\epsilon}_t' C(X_t\underline{b}_0 + \tilde{\epsilon}_t))^2 \mid \tilde{g}_t]$$

$$= E[\operatorname{tr}(C)^2 - \frac{2\operatorname{tr}(C)}{\sigma_0^2}\tilde{\epsilon}_t' C(X_t\underline{b}_0 + \tilde{\epsilon}_t) \mid \tilde{g}_t] + E[\frac{1}{\sigma_0^4}\tilde{\epsilon}_t' C(X_t\underline{b}_0 + \tilde{\epsilon}_t)(X_t\underline{b}_0 + \tilde{\epsilon}_t)' C'\tilde{\epsilon}_t \mid \tilde{g}_t]$$

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

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

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

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

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

$$= -E[\frac{\partial^2 l(\theta_0)}{\partial\rho^2} \mid \tilde{g}_t]. \text{ (by (106))}$$

By (103),

$$E[\frac{\partial^2 l(\theta_0)}{\partial\rho\partial\underline{b}} \mid \tilde{g}_t] = -\frac{1}{\sigma_0^2}E[X_t' W(I_n - \rho_0 W)^{-1}(X_t\underline{b}_0 + \tilde{\epsilon}_t) \mid \tilde{g}_t] = -\frac{1}{\sigma_0^2}X_t' C X_t\underline{b}_0. \quad (107)$$

By (100) and (101),

$$E[\frac{\partial l(\theta_0)}{\partial\rho}\frac{\partial l(\theta_0)}{\partial\underline{b}} \mid \tilde{g}_t] = -\frac{\operatorname{tr}(C)}{\sigma_0^2}E[X_t'\tilde{\epsilon}_t \mid \tilde{g}_t] + \frac{1}{\sigma_0^4}E[\tilde{\epsilon}_t' W\tilde{y}_t X_t'\tilde{\epsilon}_t \mid \tilde{g}_t]$$

$$= \frac{1}{\sigma_0^4}E[\tilde{\epsilon}_t' C(X_t\underline{b}_0 + \tilde{\epsilon}_t)X_t'\tilde{\epsilon}_t \mid \tilde{g}_t] = \frac{1}{\sigma_0^4}E[X_t'\tilde{\epsilon}_t\tilde{\epsilon}_t' C X_t\underline{b}_0 \mid \tilde{g}_t] + \frac{1}{\sigma_0^4}E[\tilde{\epsilon}_t' C\tilde{\epsilon}_t X_t'\tilde{\epsilon}_t \mid \tilde{g}_t]$$

$$= \frac{1}{\sigma_0^2}X_t' C X_t\underline{b}_0 = -E[\frac{\partial^2 l(\theta_0)}{\partial\rho\partial\underline{b}} \mid \tilde{g}_t]. \text{ (by (107))}$$

By (103),

$$E[\frac{\partial^2 l(\theta_0)}{\partial\rho\partial\sigma^2} \mid \tilde{g}_t] = -\frac{1}{\sigma_0^4}E[\tilde{\epsilon}_t' C(X_t\underline{b}_0 + \tilde{\epsilon}_t) \mid \tilde{g}_t] = -\frac{1}{\sigma_0^4}E[\tilde{\epsilon}_t' C\tilde{\epsilon}_t] = -\frac{1}{\sigma_0^4}\operatorname{tr}(E[\tilde{\epsilon}_t\tilde{\epsilon}_t' C])$$

$$= -\frac{1}{\sigma_0^2}\operatorname{tr}(C). \quad (108)$$

By (100) and (101),

$$
\begin{aligned}
&E[\frac{\partial l(\theta_0)}{\partial \rho}\frac{\partial l(\theta_0)}{\partial \sigma^2} \mid \tilde{g}_t] \\
=&\frac{n\,\mathrm{tr}(C)}{2\sigma_0^2} - \frac{\mathrm{tr}(C)}{2\sigma_0^4}E[\tilde{\epsilon}_t'\tilde{\epsilon}_t] - \frac{n}{2\sigma_0^4}E[\tilde{\epsilon}_t'W\tilde{y}_t \mid \tilde{g}_t] + \frac{1}{2\sigma_0^6}E[\tilde{\epsilon}_t'W\tilde{y}_t\tilde{\epsilon}_t'\tilde{\epsilon}_t \mid \tilde{g}_t] \\
=&-\frac{n}{2\sigma_0^4}E[\tilde{\epsilon}_t'C\tilde{\epsilon}_t] + \frac{1}{2\sigma_0^6}E[\tilde{\epsilon}_t'C(X_t\underline{b}_0 + \tilde{\epsilon}_t)\tilde{\epsilon}_t'\tilde{\epsilon}_t \mid \tilde{g}_t] = -\frac{n}{2\sigma_0^2}\,\mathrm{tr}(C) + \frac{1}{2\sigma_0^6}\,\mathrm{tr}(CE[\tilde{\epsilon}_t\tilde{\epsilon}_t'\tilde{\epsilon}_t\tilde{\epsilon}_t']) \\
=&-\frac{n}{2\sigma_0^2}\,\mathrm{tr}(C) + \frac{1}{2\sigma_0^6}(n+2)\sigma_0^4\,\mathrm{tr}(C) = \frac{1}{\sigma_0^2}\,\mathrm{tr}(C) = -E[\frac{\partial^2 l(\theta_0)}{\partial\rho\partial\sigma^2} \mid \tilde{g}_t]. \text{ (by (108))}
\end{aligned} \tag{109}
$$

By (101),

$$
E[\frac{\partial l(\theta_0)}{\partial \underline{b}}\frac{\partial l(\theta_0)}{\partial \underline{b}'} \mid \tilde{g}_t] = \frac{1}{\sigma_0^4}E[X_t'\tilde{\epsilon}_t\tilde{\epsilon}_t'X_t \mid \tilde{g}_t] = \frac{1}{\sigma_0^2}X_t'X_t = -E[\frac{\partial^2 l(\theta_0)}{\partial \underline{b}\partial \underline{b}'} \mid \tilde{g}_t],
$$

where the last equality follows from (103). By (101),

$$
\begin{aligned}
&E[\frac{\partial l(\theta_0)}{\partial \sigma^2}\frac{\partial l(\theta_0)}{\partial \underline{b}} \mid \tilde{g}_t] = -\frac{n}{2\sigma_0^4}E[X_t'\tilde{\epsilon}_t \mid \tilde{g}_t] + \frac{1}{2\sigma_0^6}E[\tilde{\epsilon}_t'\tilde{\epsilon}_tX_t'\tilde{\epsilon}_t \mid \tilde{g}_t] \\
=&-\frac{n}{2\sigma_0^4}X_t'E[\tilde{\epsilon}_t \mid \tilde{g}_t] + \frac{1}{2\sigma_0^6}E[X_t'\tilde{\epsilon}_t\tilde{\epsilon}_t'\tilde{\epsilon}_t \mid \tilde{g}_t] = \frac{1}{2\sigma_0^6}X_t'E[\tilde{\epsilon}_t\tilde{\epsilon}_t'\tilde{\epsilon}_t] \\
=&\,0 = -E[\frac{\partial^2 l(\theta_0)}{\partial\sigma^2\partial \underline{b}} \mid \tilde{g}_t],
\end{aligned}
$$

where the last equality follows from (104). At last, by (101),

$$
\begin{aligned}
&E[\left(\frac{\partial l(\theta_0)}{\partial \sigma^2}\right)^2 \mid \tilde{g}_t] = E[\left(-\frac{n}{2\sigma_0^2} + \frac{1}{2\sigma_0^4}\tilde{\epsilon}_t'\tilde{\epsilon}_t\right)^2 \mid \tilde{g}_t] \\
=&\,E[\frac{n^2}{4\sigma_0^4} + \frac{1}{4\sigma_0^8}\tilde{\epsilon}_t'\tilde{\epsilon}_t\tilde{\epsilon}_t'\tilde{\epsilon}_t - \frac{n}{2\sigma_0^6}\tilde{\epsilon}_t'\tilde{\epsilon}_t] = \frac{n}{2\sigma_0^4} = -E[\frac{\partial^2 l(\theta_0)}{\partial\sigma^2\partial\sigma^2} \mid \tilde{g}_t],
\end{aligned}
$$

where the last equality follows from (104). Hence, we have shown that $-E[H(\tilde{y}_t, \tilde{g}_t; \theta_0) \mid \tilde{g}_t] = E[s(\tilde{y}_t, \tilde{g}_t; \theta_0)s(\tilde{y}_t, \tilde{g}_t; \theta_0)' \mid \tilde{g}_t]$, which completes the proof of (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

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

where the last equality follows from (vi). This implies that $a's(\tilde{y}_t, \tilde{g}_t; \theta_0) = 0, a.s.$ 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 (105). Then, we have

$$0 = a's(\tilde{y}_t, \tilde{g}_t; \theta_0) = - a_1 \operatorname{tr}(C) - \frac{na_3}{2\sigma_0^2} + \frac{a_1}{\sigma_0^2}\tilde{\epsilon}_t' C\bar{\alpha} + \frac{a_1}{\sigma_0^2}\tilde{\epsilon}_t' CB_0\tilde{g}_t + \frac{a_1}{\sigma_0^2}\tilde{\epsilon}_t' C\tilde{\epsilon}_t \\ + \frac{1}{\sigma_0^2}v'\tilde{\epsilon}_t + \frac{1}{\sigma_0^2}\tilde{\epsilon}_t' U\tilde{g}_t + \frac{a_3}{2\sigma_0^4}\tilde{\epsilon}_t'\tilde{\epsilon}_t, \ a.s. \tag{110}$$

It follows from (110) and part (ii) of Assumption D.1 that

$$0 = - a_1 \operatorname{tr}(C) - \frac{na_3}{2\sigma_0^2} + \frac{a_1}{\sigma_0^2}\tilde{\epsilon}' C\bar{\alpha}_0 + \frac{a_1}{\sigma_0^2}\tilde{\epsilon}' CB\tilde{g} + \frac{a_1}{\sigma_0^2}\tilde{\epsilon}' C\tilde{\epsilon} \\ + \frac{1}{\sigma_0^2}v'\tilde{\epsilon} + \frac{1}{\sigma_0^2}\tilde{\epsilon}' U\tilde{g} + \frac{a_3}{2\sigma_0^4}\tilde{\epsilon}'\tilde{\epsilon}, \text{ for any } (\tilde{\epsilon}', \tilde{g}')' \in \mathbb{R}^n \times \mathbb{A}. \tag{111}$$

By taking partial derivatives with respect to $(\tilde{\epsilon}', \tilde{g}')'$ on both sides of (111), we obtain that

$$- a_1 \operatorname{tr}(C) - \frac{na_3}{2\sigma_0^2} = 0, \ \frac{a_1}{\sigma_0^2}C\bar{\alpha}_0 + \frac{1}{\sigma_0^2}v = 0, \ \frac{a_1}{\sigma_0^2}CB_0 + \frac{1}{\sigma_0^2}U = 0, \tag{112}$$

$$\frac{a_1}{\sigma_0^2}C + \frac{a_3}{2\sigma_0^4}I_n = 0. \tag{113}$$

There are two cases:

Case 1: $a_1 = 0$. Then, it follows from (112) 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 (113) that $C = -\frac{a_3}{2\sigma_0^2 a_1}I_n$, which in combination with (105) 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 D.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 D.2). Hence, $E[H(\tilde{y}_t, \tilde{g}_t; \theta_0)]$ is invertible.

(viii) Let $\mathcal{N}$ be any neighborhood of θ_0 that lies in the interior of Θ. We have

$$E[\sup_{\theta\in\mathcal{N}} \|H(\tilde{y}_t, \tilde{g}_t; \theta)\|] \leq E[\sup_{\theta\in\mathcal{N}} |\frac{\partial^2 l(\theta)}{\partial\rho^2}|] + E[\sup_{\theta\in\mathcal{N}} \|\frac{\partial^2 l(\theta)}{\partial\rho\partial\underline{b}}\|] + E[\sup_{\theta\in\mathcal{N}} |\frac{\partial^2 l(\theta)}{\partial\rho\partial\sigma_2}|] \\ + E[\sup_{\theta\in\mathcal{N}} \|\frac{\partial^2 l(\theta)}{\partial\underline{b}\partial\underline{b}'}\|] + E[\sup_{\theta\in\mathcal{N}} \|\frac{\partial^2 l(\theta)}{\partial\sigma^2\partial\underline{b}}\|] + E[\sup_{\theta\in\mathcal{N}} |\frac{\partial^2 l(\theta)}{\partial\sigma^2\partial\sigma^2}|\|]. \tag{114}$$

We only need to show that each term on the right side of (114) is finite. Let $C(\rho) := W(I_n - \rho W)^{-1}$, then $C(\rho)$ is continuous. By (102), (103), and (104), we have

$$E[\sup_{\theta\in\mathcal{N}} |\frac{\partial^2 l(\theta)}{\partial\rho^2}|] \leq \sup_{\theta\in\mathcal{N}} |\operatorname{tr}(C(\rho)^2)| + E[\tilde{y}_t' W' W\tilde{y}_t] \sup_{\theta\in\mathcal{N}} \frac{1}{\sigma^2}, \tag{115}$$

$$E[\sup_{\theta\in\mathcal{N}}\|\frac{\partial^2 l(\theta)}{\partial\rho\partial\underline{b}}\|] \leq E[\|X_t'W\tilde{y}_t\|]\sup_{\theta\in\mathcal{N}}\frac{1}{\sigma^2}, \tag{116}$$

$$\begin{aligned}
&E[\sup_{\theta\in\mathcal{N}}|\frac{\partial^2 l(\theta)}{\partial\rho\partial\sigma^2}|] = E[\sup_{\theta\in\mathcal{N}}|\frac{1}{\sigma^4}\tilde{\xi}_t'W\tilde{y}_t|] \\
&\leq E[\sup_{\theta\in\mathcal{N}}|\frac{1}{\sigma^4}\tilde{y}_t'W\tilde{y}_t|] + E[\sup_{\theta\in\mathcal{N}}|\frac{\rho}{\sigma^4}\tilde{y}_t'W'W\tilde{y}_t|] + E[\sup_{\theta\in\mathcal{N}}|\frac{1}{\sigma^4}\bar{\alpha}'W\tilde{y}_t|] + E[\sup_{\theta\in\mathcal{N}}|\frac{1}{\sigma^4}\tilde{g}_t'B'W\tilde{y}_t|] \\
&\leq E[|\tilde{y}_t'W\tilde{y}_t|]\sup_{\theta\in\mathcal{N}}\frac{1}{\sigma^4} + E[\tilde{y}_t'W'W\tilde{y}_t]\sup_{\theta\in\mathcal{N}}\frac{|\rho|}{\sigma^4} + E[\|W\tilde{y}_t\|]\sup_{\theta\in\mathcal{N}}\frac{1}{\sigma^4}\|\bar{\alpha}\| \\
&\quad + E[\|\tilde{g}_t\|^2]\sup_{\theta\in\mathcal{N}}\frac{1}{2\sigma^4}\|B\|^2 + E[\|W\tilde{y}_t\|^2]\sup_{\theta\in\mathcal{N}}\frac{1}{2\sigma^4},
\end{aligned} \tag{117}$$

$$E[\sup_{\theta\in\mathcal{N}}\|\frac{\partial^2 l(\theta)}{\partial\underline{b}\partial\underline{b}'}\|] \leq E[\|X_t'X_t\|]\sup_{\theta\in\mathcal{N}}\frac{1}{\sigma^2}, \tag{118}$$

$$\begin{aligned}
&E[\sup_{\theta\in\mathcal{N}}\|\frac{\partial^2 l(\theta)}{\partial\sigma^2\partial\underline{b}}\|] = E[\sup_{\theta\in\mathcal{N}}\|\frac{1}{\sigma^4}X_t'\tilde{\xi}_t\|] \\
&\leq E[\sup_{\theta\in\mathcal{N}}\|\frac{1}{\sigma^4}X_t'\tilde{y}_t\|] + E[\sup_{\theta\in\mathcal{N}}\|\frac{\rho}{\sigma^4}X_t'W\tilde{y}_t\|] + E[\sup_{\theta\in\mathcal{N}}\|\frac{1}{\sigma^4}X_t'\bar{\alpha}\|] + E[\sup_{\theta\in\mathcal{N}}\|\frac{1}{\sigma^4}X_t'B\tilde{g}_t\|] \\
&\leq E[\|X_t'\tilde{y}_t\|]\sup_{\theta\in\mathcal{N}}\frac{1}{\sigma^4} + E[\|X_t'W\tilde{y}_t\|]\sup_{\theta\in\mathcal{N}}\frac{|\rho|}{\sigma^4} + E[\|X_t\|]\sup_{\theta\in\mathcal{N}}\frac{\|\bar{\alpha}\|}{\sigma^4} + E[\|X_t'\|\|\tilde{g}_t\|]\sup_{\theta\in\mathcal{N}}\frac{\|B\|}{\sigma^4} \\
&\leq E[\|X_t'\tilde{y}_t\|]\sup_{\theta\in\mathcal{N}}\frac{1}{\sigma^4} + E[\|X_t'W\tilde{y}_t\|]\sup_{\theta\in\mathcal{N}}\frac{|\rho|}{\sigma^4} + E[\|X_t\|]\sup_{\theta\in\mathcal{N}}\frac{\|\bar{\alpha}\|}{\sigma^4} \\
&\quad + \frac{1}{2}(E[\|X_t'\|^2] + E[\|\tilde{g}_t\|^2])\sup_{\theta\in\mathcal{N}}\frac{\|B\|}{\sigma^4},
\end{aligned} \tag{119}$$

$$\begin{aligned}
&E[\sup_{\theta\in\mathcal{N}}|\frac{\partial^2 l(\theta)}{\partial\sigma^2\partial\sigma^2}|] = E[\sup_{\theta\in\mathcal{N}}|\frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\tilde{\xi}_t'\tilde{\xi}_t|] \\
&= E[\sup_{\theta\in\mathcal{N}}|\frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\|(I_n - \rho W)\tilde{y}_t - \bar{\alpha} - B\tilde{g}_t\|^2|] \\
&\leq E[\sup_{\theta\in\mathcal{N}}(\frac{n}{2\sigma^4} + \frac{16}{\sigma^6}(\|\tilde{y}_t\|^2 + \rho^2\|W\tilde{y}_t\|^2 + \|\bar{\alpha}\|^2 + \|B\|^2\|\tilde{g}_t\|^2))] \\
&\leq \sup_{\theta\in\mathcal{N}}(\frac{n}{2\sigma^4} + \frac{16\|\bar{\alpha}\|^2}{\sigma^6}) + E[\|\tilde{y}_t\|^2]\sup_{\theta\in\mathcal{N}}\frac{16}{\sigma^6} + E[\|W\tilde{y}_t\|^2]\sup_{\theta\in\mathcal{N}}\frac{16\rho^2}{\sigma^6} + E[\|\tilde{g}_t\|^2]\sup_{\theta\in\mathcal{N}}\frac{16\|B\|^2}{\sigma^6}.
\end{aligned} \tag{120}$$

Since Θ is compact, all the supremums on the right-hand side of (115)-(120) are finite. Furthermore, by part (i) of Assumption D.1, $\tilde{g}_t$ and hence $\tilde{y}_t$ have finite second moments. Thus, all the expectations on the right-hand side of (115)-(120) are finite. Therefore, each term on the right-hand side of (114) is finite, which completes the proof. ∎

Theorem D.1. *(Asymptotic properties of the MLE) The MLE* $\hat{\theta} := (\hat{\rho}, \hat{\underline{b}}, \widehat{\sigma^2})$ *has* *consistency and asymptotic normality:*

$$(i)\ \hat{\theta} \xrightarrow{p} \theta_0,\ \text{as } T \to \infty, \tag{121}$$

$$(ii)\ \sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, -E[H(\tilde{y}_t, \tilde{g}_t; \theta_0)]^{-1}),\ \text{as } T \to \infty, \tag{122}$$

$$(iii)\ \frac{1}{T}\sum_{t=1}^{T} H(\tilde{y}_t, \tilde{g}_t; \hat{\theta}) \xrightarrow{p} E[H(\tilde{y}_t, \tilde{g}_t; \theta_0)],\ \text{as } T \to \infty, \tag{123}$$

where $H(\tilde{y}_t, \tilde{g}_t; \theta)$ is equal to (97).

Proof. By (i), (ii), and (iii) of Proposition D.2 and the compactness of Θ, it follows from Theorem 2.1 in Neway and McFadden (1994) that (121) holds.

We will show the asymptotic normality (122) by applying Proposition 7.9 in Hayashi (2000, p. 475). The consistency of $\hat{\theta}$ has been proved in above. The condition (1) of Proposition 7.9 holds by the assumption that θ_0 lies in the interior of Θ. The conditions (2), (3), (4), and (5) of Proposition 7.9 follow from Proposition D.2 of this paper. Hence, all the conditions of Proposition 7.9 hold and its conclusion implies (122).

At last, we will show that (123) holds. Define $\hat{H}(\theta) := \frac{1}{T}\sum_{t=1}^{T} H(\tilde{y}_t, \tilde{g}_t; \theta)$. Let $\mathcal{N}$ be a neighborhood such that (viii) in Proposition D.2 holds and let $\Theta_0 \subset \mathcal{N}$ be a compact set that contains θ_0. Then, it follows from (viii) in Proposition D.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

$$\sup_{\theta \in \Theta_0} \left\| \frac{1}{T}\sum_{t=1}^{T} \frac{\partial^2 l(\tilde{y}_t \mid \tilde{g}_t, \theta)}{\partial\theta\partial\theta'} - H(\theta) \right\| \xrightarrow{p} 0,\ \text{as } T \to \infty. \tag{124}$$

Since $\hat{\theta} \xrightarrow{p} \theta_0$ and $H(\theta)$ is continuous, it follows that $H(\hat{\theta}) \xrightarrow{p} H(\theta_0)$. For any $\varepsilon > 0$,

$$\begin{aligned}
&P(\|\hat{H}(\hat{\theta}) - H(\theta_0)\| > \varepsilon) \leq P(\|\hat{H}(\hat{\theta}) - H(\hat{\theta})\| + \|H(\hat{\theta}) - H(\theta_0)\| > \varepsilon) \\
&\leq P(\|\hat{H}(\hat{\theta}) - H(\hat{\theta})\| > \frac{\varepsilon}{2}, \hat{\theta} \in \Theta_0) + P(\|\hat{H}(\hat{\theta}) - H(\hat{\theta})\| > \frac{\varepsilon}{2}, \hat{\theta} \notin \Theta_0) \\
&\quad + P(\|H(\hat{\theta}) - H(\theta_0)\| > \frac{\varepsilon}{2}) \\
&\leq P(\sup_{\theta \in \Theta_0} \|\hat{H}(\theta) - H(\theta)\| > \frac{\varepsilon}{2}) + P(\hat{\theta} \notin \Theta_0) + P(\|H(\hat{\theta}) - H(\theta_0)\| > \frac{\varepsilon}{2}) \\
&\to 0,\ \text{as } T \to \infty,
\end{aligned}$$

where the limit follows from (124), $\hat{\theta} \xrightarrow{p} \theta_0$, and $H(\hat{\theta}) \xrightarrow{p} H(\theta_0)$. ∎

The asymptotic properties of the MLE of S-APT model (33) 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 D.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).[18]

We investigate the finite-sample performance of the estimators using 2000 data sets simulated from the model (33). In all the simulation studies, 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.[19] It is easy to check that the specified matrix W is regular;[20] hence, by Proposition D.1, the true parameter is 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 (33). Then, the MLE $\hat{\rho}$ is obtained from each of the simulated data sets.

Table 7 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 4 shows the histogram of 2000 estimates $\hat{\rho}$ for different ρ_0, which seems to indicate that $\hat{\rho}$ has an asymptotic normal distribution with mean ρ_0.

D.3 Simulation Studies for Testing S-APT and Estimating Adjusted R^2

Using the likelihood ratio test statistic defined in (40), we test the zero-intercept constraint of S-APT for 10000 simulated data sets at the confidence level of 95%.[21]

[18] 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.

[19] 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.

[20] This is because the sum of square of different columns of W are not equal.

[21] The test data include the 8000 data sets used for Table 7 and additional 2000 data sets simulated in the same way with $\rho_0 = 0.5$.

ρ_0	0.2	0.4	0.6	0.8
mean of $\hat{\rho}$	0.199	0.399	0.599	0.799
(theoretical) asymptotic standard deviation of $\hat{\rho}$	0.028	0.025	0.019	0.011
empirical standard deviation of $\hat{\rho}$	0.031	0.029	0.020	0.013

Table 7: The mean and standard deviation of $\hat{\rho}$. The asymptotic standard deviations are estimated from the sample average of Hessian matrix (see Eq. (97) and (123)).

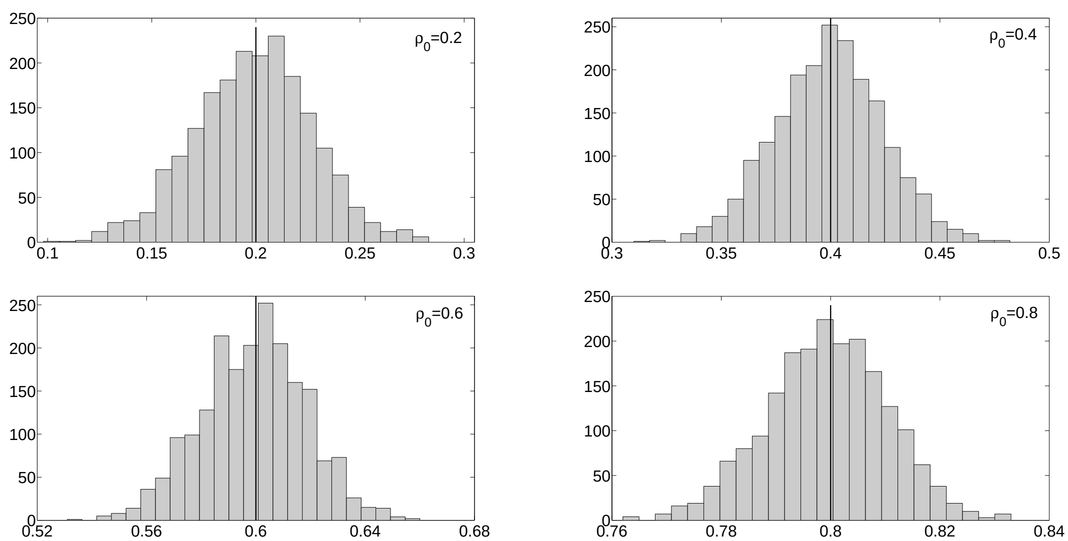


Figure 4: Histogram of the MLE $\hat{\rho}$ for 2000 data sets simulated from the model (33) for different values of ρ_0 with $n = 20$, $K = 1$, and $T = 131$.

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).

To show the effectiveness of the adjusted R^2, two data sets are simulated according to the same model specification as that in Table 7, except that ρ_0 is fixed at 0.5, and two values of σ_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, the factor realization $\{\tilde{g}_t, t = 1, \ldots, T\}$ is first simulated and fixed. Then, for each chosen value of σ_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 calculated according to the model (33). For each simulated data set, we calculate the MLE estimate θ^* under the constraint $\bar{\alpha}_0 = 0$ and obtain the fitted residual series $\{\hat{\epsilon}_t = \tilde{y}_t - \rho^* W \tilde{y}_t - B^* \tilde{g}_t : t = 1, \ldots, T\}$, where ρ^* and B^* are the MLE. The sample adjusted R^2 of y_i is computed and compared to the theoretical adjusted R^2 of y_i. Table 8 shows that the sample adjusted R^2 and the theoretical adjusted R^2 align well.

	$\sigma_0^2 = 0.01$									
	r_1	r_2	r_3	r_4	r_5	r_6	r_7	r_8	r_9	r_{10}
theoretical adjusted R^2	0.7763	0.9818	0.0910	0.4064	0.9654	0.9730	0.9690	0.3210	0.8415	0.3110
sample adjusted R^2	0.7748	0.9817	0.0848	0.4023	0.9652	0.9728	0.9688	0.3162	0.8404	0.3062
	r_{11}	r_{12}	r_{13}	r_{14}	r_{15}	r_{16}	r_{17}	r_{18}	r_{19}	r_{20}
theoretical adjusted R^2	0.1833	0.9203	0.8952	0.9906	0.6367	0.6634	0.9672	0.0451	0.2236	0.9394
sample adjusted R^2	0.1777	0.9197	0.8945	0.9905	0.6342	0.6611	0.9670	0.0385	0.2183	0.9390
	$\sigma_0^2 = 0.5$									
	r_1	r_2	r_3	r_4	r_5	r_6	r_7	r_8	r_9	r_{10}
theoretical adjusted R^2	0.2328	0.4256	0.0007	0.1308	0.4691	0.3521	0.4452	0.0859	0.0567	0.0002
sample adjusted R^2	0.2294	0.4230	-0.0037	0.1269	0.4667	0.3491	0.4427	0.0817	0.0525	-0.0043
	r_{11}	r_{12}	r_{13}	r_{14}	r_{15}	r_{16}	r_{17}	r_{18}	r_{19}	r_{20}
theoretical adjusted R^2	0.1003	0.2366	0.1816	0.6552	0.1691	0.0235	0.4598	-0.0705	0.0390	0.1301
sample adjusted R^2	0.0961	0.2331	0.1779	0.6536	0.1654	0.0191	0.4574	-0.0754	0.0347	0.1262

Table 8: Simulation study of the sample adjusted R^2. We use the same model specification as that for Table 7, except that ρ_0 is fixed at 0.5 and two values of σ_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 (33) under the constraint $\bar{\alpha}_0 = 0$ 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.

E Spatial Weight Matrices and S&P Credit Ratings

The spatial weight matrix W that is used in Section 6 and calculated using driving distance is

$$\begin{pmatrix}
0.000 & 0.080 & 0.082 & 0.116 & 0.210 & 0.083 & 0.069 & 0.127 & 0.125 & 0.049 & 0.059 \\
0.041 & 0.000 & 0.036 & 0.236 & 0.095 & 0.026 & 0.075 & 0.050 & 0.359 & 0.036 & 0.046 \\
0.132 & 0.113 & 0.000 & 0.147 & 0.142 & 0.067 & 0.077 & 0.080 & 0.121 & 0.057 & 0.064 \\
0.074 & 0.295 & 0.058 & 0.000 & 0.087 & 0.031 & 0.086 & 0.064 & 0.182 & 0.052 & 0.072 \\
0.176 & 0.157 & 0.074 & 0.114 & 0.000 & 0.051 & 0.070 & 0.079 & 0.183 & 0.043 & 0.052 \\
0.146 & 0.090 & 0.073 & 0.086 & 0.107 & 0.000 & 0.066 & 0.198 & 0.088 & 0.067 & 0.078 \\
0.084 & 0.179 & 0.059 & 0.164 & 0.102 & 0.046 & 0.000 & 0.074 & 0.154 & 0.063 & 0.075 \\
0.148 & 0.114 & 0.058 & 0.117 & 0.110 & 0.131 & 0.071 & 0.000 & 0.100 & 0.066 & 0.085 \\
0.069 & 0.388 & 0.041 & 0.157 & 0.120 & 0.027 & 0.070 & 0.047 & 0.000 & 0.035 & 0.045 \\
0.068 & 0.097 & 0.048 & 0.113 & 0.071 & 0.052 & 0.071 & 0.078 & 0.088 & 0.000 & 0.315 \\
0.069 & 0.107 & 0.046 & 0.132 & 0.072 & 0.052 & 0.073 & 0.085 & 0.095 & 0.268 & 0.000
\end{pmatrix}$$

The spatial weight matrix W that is used in Section 6 and calculated using geographic distance is

$$\begin{pmatrix}
0.000 & 0.116 & 0.074 & 0.103 & 0.203 & 0.083 & 0.063 & 0.139 & 0.113 & 0.046 & 0.059 \\
0.065 & 0.000 & 0.036 & 0.226 & 0.092 & 0.029 & 0.077 & 0.051 & 0.344 & 0.035 & 0.045 \\
0.129 & 0.113 & 0.000 & 0.097 & 0.168 & 0.075 & 0.092 & 0.084 & 0.124 & 0.055 & 0.063 \\
0.074 & 0.291 & 0.040 & 0.000 & 0.087 & 0.037 & 0.098 & 0.069 & 0.178 & 0.053 & 0.073 \\
0.184 & 0.149 & 0.087 & 0.110 & 0.000 & 0.054 & 0.073 & 0.082 & 0.168 & 0.042 & 0.052 \\
0.149 & 0.091 & 0.077 & 0.091 & 0.106 & 0.000 & 0.067 & 0.182 & 0.088 & 0.067 & 0.081 \\
0.075 & 0.163 & 0.062 & 0.162 & 0.096 & 0.044 & 0.000 & 0.067 & 0.167 & 0.077 & 0.087 \\
0.166 & 0.108 & 0.058 & 0.115 & 0.107 & 0.121 & 0.067 & 0.000 & 0.098 & 0.068 & 0.093 \\
0.070 & 0.376 & 0.043 & 0.152 & 0.113 & 0.030 & 0.086 & 0.050 & 0.000 & 0.035 & 0.044 \\
0.067 & 0.090 & 0.046 & 0.107 & 0.067 & 0.054 & 0.095 & 0.083 & 0.083 & 0.000 & 0.307 \\
0.072 & 0.099 & 0.044 & 0.124 & 0.070 & 0.055 & 0.090 & 0.096 & 0.088 & 0.260 & 0.000
\end{pmatrix}.$$

The spatial weight matrix for 10 CSI indices futures that is used in Section 7 and calculated using driving distance is

$$\begin{pmatrix}
0.000 & 0.468 & 0.152 & 0.057 & 0.022 & 0.021 & 0.029 & 0.019 & 0.211 & 0.021 \\
0.509 & 0.000 & 0.125 & 0.058 & 0.023 & 0.024 & 0.030 & 0.021 & 0.188 & 0.022 \\
0.293 & 0.221 & 0.000 & 0.088 & 0.039 & 0.037 & 0.052 & 0.036 & 0.196 & 0.038 \\
0.137 & 0.129 & 0.110 & 0.000 & 0.083 & 0.067 & 0.139 & 0.071 & 0.187 & 0.078 \\
0.034 & 0.034 & 0.032 & 0.054 & 0.000 & 0.088 & 0.129 & 0.207 & 0.037 & 0.385 \\
0.072 & 0.074 & 0.064 & 0.095 & 0.189 & 0.000 & 0.143 & 0.133 & 0.077 & 0.154 \\
0.067 & 0.065 & 0.063 & 0.134 & 0.192 & 0.099 & 0.000 & 0.137 & 0.077 & 0.167 \\
0.032 & 0.031 & 0.030 & 0.048 & 0.215 & 0.064 & 0.096 & 0.000 & 0.035 & 0.449 \\
0.307 & 0.252 & 0.148 & 0.113 & 0.034 & 0.033 & 0.048 & 0.031 & 0.000 & 0.033 \\
0.027 & 0.027 & 0.026 & 0.043 & 0.325 & 0.060 & 0.095 & 0.366 & 0.030 & 0.000
\end{pmatrix}.$$

The spatial weight matrix W for 10 CSI indices futures that is used in Section 7 and calculated using geographic distance is

$$\begin{pmatrix}
0.000 & 0.456 & 0.147 & 0.061 & 0.022 & 0.022 & 0.029 & 0.020 & 0.222 & 0.021 \\
0.488 & 0.000 & 0.119 & 0.065 & 0.024 & 0.024 & 0.031 & 0.021 & 0.205 & 0.022 \\
0.272 & 0.206 & 0.000 & 0.099 & 0.039 & 0.036 & 0.051 & 0.035 & 0.226 & 0.037 \\
0.140 & 0.139 & 0.122 & 0.000 & 0.078 & 0.067 & 0.126 & 0.066 & 0.191 & 0.071 \\
0.034 & 0.035 & 0.032 & 0.053 & 0.000 & 0.086 & 0.133 & 0.201 & 0.038 & 0.388 \\
0.072 & 0.074 & 0.065 & 0.098 & 0.183 & 0.000 & 0.142 & 0.134 & 0.077 & 0.155 \\
0.067 & 0.068 & 0.063 & 0.128 & 0.197 & 0.099 & 0.000 & 0.138 & 0.077 & 0.164 \\
0.032 & 0.033 & 0.031 & 0.048 & 0.213 & 0.067 & 0.099 & 0.000 & 0.035 & 0.442 \\
0.298 & 0.258 & 0.163 & 0.113 & 0.033 & 0.031 & 0.045 & 0.029 & 0.000 & 0.031 \\
0.028 & 0.028 & 0.026 & 0.042 & 0.332 & 0.062 & 0.095 & 0.357 & 0.030 & 0.000
\end{pmatrix}.$$

The spatial weight matrix W for 9 CSI indices futures (excluding New York) that is used in Section 7 and calculated using driving distance is

$$\begin{pmatrix} 0.000 & 0.478 & 0.156 & 0.058 & 0.022 & 0.022 & 0.029 & 0.020 & 0.216 \\ 0.520 & 0.000 & 0.128 & 0.060 & 0.024 & 0.024 & 0.031 & 0.021 & 0.193 \\ 0.305 & 0.230 & 0.000 & 0.091 & 0.041 & 0.038 & 0.054 & 0.037 & 0.204 \\ 0.148 & 0.140 & 0.119 & 0.000 & 0.090 & 0.073 & 0.150 & 0.077 & 0.203 \\ 0.055 & 0.055 & 0.052 & 0.088 & 0.000 & 0.142 & 0.211 & 0.337 & 0.060 \\ 0.085 & 0.087 & 0.076 & 0.112 & 0.223 & 0.000 & 0.169 & 0.157 & 0.092 \\ 0.080 & 0.078 & 0.076 & 0.161 & 0.230 & 0.118 & 0.000 & 0.165 & 0.093 \\ 0.057 & 0.056 & 0.055 & 0.087 & 0.390 & 0.116 & 0.175 & 0.000 & 0.063 \\ 0.318 & 0.261 & 0.153 & 0.117 & 0.035 & 0.034 & 0.050 & 0.032 & 0.000 \end{pmatrix}.$$

The spatial weight matrix W for 9 CSI indices futures (excluding New York) that is used in Section 7 and calculated using geographic distance is

$$\begin{pmatrix} 0.000 & 0.466 & 0.150 & 0.063 & 0.023 & 0.022 & 0.030 & 0.020 & 0.227 \\ 0.499 & 0.000 & 0.121 & 0.067 & 0.024 & 0.025 & 0.032 & 0.022 & 0.210 \\ 0.282 & 0.214 & 0.000 & 0.103 & 0.040 & 0.038 & 0.053 & 0.036 & 0.234 \\ 0.150 & 0.150 & 0.131 & 0.000 & 0.084 & 0.072 & 0.136 & 0.071 & 0.206 \\ 0.056 & 0.057 & 0.053 & 0.087 & 0.000 & 0.140 & 0.217 & 0.328 & 0.062 \\ 0.085 & 0.088 & 0.077 & 0.116 & 0.216 & 0.000 & 0.168 & 0.159 & 0.091 \\ 0.080 & 0.081 & 0.075 & 0.153 & 0.236 & 0.118 & 0.000 & 0.165 & 0.092 \\ 0.058 & 0.058 & 0.056 & 0.085 & 0.382 & 0.120 & 0.177 & 0.000 & 0.063 \\ 0.308 & 0.266 & 0.169 & 0.116 & 0.034 & 0.032 & 0.046 & 0.030 & 0.000 \end{pmatrix}.$$

	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Austria	AAA	AAA	AAA	AAA	AAA	AAA	AAA	AAA	AAA	AAA	AAA	AA+	AA+
Belgium	AA+	AA+	AA+	AA+	AA+	AA+	AA+	AA+	AA+	AA+	AA	AA	AA
Finland	AA+	AAA	AAA	AAA	AAA	AAA	AAA	AAA	AAA	AAA	AAA	AAA	AAA
France	AAA	AAA	AAA	AAA	AAA	AAA	AAA	AAA	AAA	AAA	AA+	AA+	AA+
Germany	AAA	AAA	AAA	AAA	AAA	AAA	AAA	AAA	AAA	AAA	AAA	AAA	AAA
Greece	A	A	A+	A+	A	A	A	A	A-	BBB+/BB+	BB/B/CCC/CC	SD/CCC	B-
Ireland	AAA	AAA	AAA	AAA	AAA	AAA	AAA	AAA	AAA	AA+/AA	AA-/A	A-/BBB+	BBB+
Italy	AA	AA	AA	AA	AA-	AA-	A+	A+	A+	A	BBB+	BBB+	
Netherlands	AAA	AAA	AAA	AAA	AAA	AAA	AAA	AAA	AAA	AAA	AAA	AAA	AA+
Portugal	AA	AA	AA	AA	AA-	AA-	AA-	AA-	A+	A-	BBB-	BB	BB
Spain	AA+	AA+	AA+	AA+	AAA	AAA	AAA	AAA	AA+	AA	AA-	A/BBB+/BBB-	BBB-

Table 9: The S&P credit ratings of the 11 Eurozone countries during 2001–2013.

	2013	2012	2011	2010	2009	2008	2007	2006
Florida	AAA	AAA	AAA	AAA	AAA	AAA	AAA	AAA
Nevada	AA	AA	AA	AA+	AA+	AA+	AA+	AA+
Massachusetts	AA+	AA+	AA	AA	AA	AA	AA	AA
New York	AA	AA	AA	AA	AA	AA	AA	AA
Colorado	AA	AA	AA	AA	AA	AA	AA	AA-
Illinois	A-	A+	A+	A+	A+	A	A	A
California	A	A-	A-	A-	A	A+	A+	A+

Table 10: The S&P credit ratings of the states where the ten MSAs are located during 2006–2013.

References

Anselin, L. 1988. *Spatial Econometrics: Methods and Models*. Kluwer, Dordrecht.

Anselin, L., J. Le Gallo, J. Jayet. 2008. Spatial panel econometrics. *The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice*. Springer, Berlin.

Bekaert, G., R. J. Hodrick, X. Y. Zhang. 2009. International stock return comovements. *Journal of Finance* **64** 2591–2626.

Black, F. 1976. The pricing of commodity contracts. *Journal of Financial Economics* **3** 167–179.

Campbell, J. Y., A. W. Lo, A. C. MacKinlay. 1996. *The Econometrics of Financial Markets*. Princeton University Press, Princeton.

Case, K. E., R. J. Shiller. 1987. Prices of single-family homes since 1970: New indices for four cities. *New England Economic Review* 45–56.

Chamberlain, G. 1983. Funds, factors and diversification in arbitrage pricing models. *Econometrica* **51** 1305–1323.

Chamberlain, G., M. Rothschild. 1983. Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica* **51** 1281–1304.

Coval, J. D., T. J. Moskowitz. 2001. The geography of investment: Informed trading and asset prices. *Journal of Political Economy* **109** 811–841.

Cressie, N. 1993. *Statistics for Spatial Data*. John Wiley, New York.

De Roon, F. A., T. E. Nijman, C. Veld. 2000. Hedging pressure effects in futures markets. *Journal of Finance* **55** 1437–1456.

Duan, J. C., J. Sun, T. Wang. 2012. Multiperiod corporate default prediction–a forward intensity approach. *Journal of Econometrics* **170** 191–209.

Duffie, D. 1989. *Futures Markets*. Prentice-Hall, New Jersey.

Fabozzi, F. J., R. J. Shiller, R. S. Tunaru. 2012. A pricing framework for real estate derivatives. *European Financial Management* **18** 762–789.

Fama, E. F., K. R. French. 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* **33** 3–56.

Fama, E. F., K. R. French. 2012. Size, value, and momentum in international stock returns. *Journal of Financial Economics* **105** 457–472.

Fernandez, V. 2011. Spatial linkages in international financial markets. *Quantitative Finance* **11** 237–245.

Gibbons, M. R., S. A. Ross, J. Shanken. 1989. A test of the efficiency of a given portfolio. *Econometrica* **57** 1121–1152.

Hayashi, F. 2000. *Econometrics*. Princeton University Press, Princeton.

Holly, S., M. H. Pesaran, T. Yamagata. 2010. A spatio-temporal model of housing prices in the USA. *Journal of Econometrics* **158** 160–173.

Huberman, G. A. 1982. A simple approach to arbitrage pricing theory. *Journal of Economic Theory* **28** 183–192.

Huberman, G. A., Z. Wang. 2008. Arbitrage pricing theory. S. N. Durlauf, L. E. Blume, eds., *The New Palgrave Dictionary of Economics*, 2nd ed.

Ingersoll, J. E. 1984. Some results in the theory of arbitrage pricing. *Journal of Finance* **39** 1021–1039.

Lee, L. F. 2004. Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica* **72** 1899–1925.

Lee, L. F., J. Yu. 2010a. Estimation of spatial autoregressive panel data models with fixed effects. *Journal of Econometrics* **154** 165–185.

Lee, L. F., J. Yu. 2010b. Some recent developments in spatial panel data models. *Regional Science and Urban Economics* **40** 255–271.

Lesage, J., R. K. Pace. 2009. *Introduction to Spatial Econometrics*. Chapman & Hall/CRC Press, Boca Raton.

Markowitz, H. 1952. Portfolio selection. *Journal of Finance* **7** 77–91.

Neway, W., D. McFadden. 1994. Large sample estimation and hypothesis testing. *Handbook of Econometrics* **4** 2111–2245.

Ord, K. 1975. Estimation methods for models of spatial interaction. *Journal of the American Statistical Association* **70** 120–126.

Pace, L. 2003. *Spatial Statistics Toolbox 2.0*. Manuscript.

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

Pirinsky, C., Q. Wang. 2006. Does corporate headquarters location matter for stock returns? *Journal of Finance* **61** 1991–2015.

Ramsey, J. B. 1969. Tests for specification errors in classical linear least squares regression analysis. *Journal of the Royal Statistical Society, Series B* **31** 350–371.

Ross, S. 1976a. The arbitrage theory of capital asset pricing. *Journal of Economic Theory* **13** 341–360.

Ross, S. 1976b. Risk, return and arbitrage. I. Friend, J. Bicksler, eds., *Risk Return in Finance*, vol. 1. Ballinger, Cambridge, 189–218.

Sharpe, W. F. 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance* **19** 425–442.

Shiller, R. J. 1993. *Macro Markets*. Clarendon Press, Oxford.

Vincenty, T. 1975. Direct and inverse solutions of geodesics on the ellipsoid with application of nested equations. *Survey Review* **176** 88–93.

Whittle, P. 1954. On stationary processes in the plane. *Biometrika* **41** 434–449.