Abnormal Returns or Mismeasured Risk? Network Effects and Risk Spillover in Stock Returns

Abnormal Returns or Mismeasured Risk? Network Effects and Risk Spillover in Stock Returns

Arnab Bhattacharjee [1] and **Sudipto Roy** [2,*]

1   Spatial Economics & Econometrics Centre, Heriot-Watt University, Edinburgh EH14 4AS, UK;
    a.bhattacharjee@hw.ac.uk
2   Finlabs India Pvt. Ltd., Mumbai 400051, India
*   Correspondence: sudipto.r@finlabsindia.com; Tel.: +44-131-451-3482

Abstract: Recent event study literature has highlighted abnormal stock returns, particularly in short event windows. A common explanation is the cross-correlation of stock returns that are often enhanced during periods of sharp market movements. This suggests the misspecification of the underlying factor model, typically the Fama-French model. By drawing upon recent panel data literature with cross-section dependence, we argue that the Fame-French factor model can be enriched by allowing explicitly for network effects between stock returns. We show that recent empirical work is consistent with the above interpretation, and we advance some hypotheses along which new structural models for stock returns may be developed. Applied to data on stock returns for the 30 Dow Jones Industrial Average (DJIA) stocks, our framework provides exciting new insights.

Keywords: Fama-French factor model; market microstructure; trading behavior; panel data factor model; social network model; risk spillover; abnormal returns

1. Introduction

In finance theory and empirics, stock returns are typically described by a factor model along the lines of Fama and French (1988, 1993, 2015) and Carhart (1997). However, despite the popularity of the Fama-French (FF) and FF-type models, substantial literature in the event study tradition, starting from Brown and Warner (1985) and Strong (1992), has pointed towards a failure of the FF model to adequately capture the relationship between risk and return; for recent literature, see Chiang and Li (2012) and Marks and Musumeci (2017), among others. Specifically, there are periods when stock returns are highly correlated (Kolari and Pynnönen 2010); this correlation leads to abnormal returns and mismeasured risk; see also Boehmer et al. (1991) and Kothari and Warner (2007).[1]

In this paper, we contrast the FF-type factor models for stock returns against the standard panel data factor model in contemporary econometrics. Then, recent developments in the econometrics of panel factor models with cross-section dependence suggest reasons why the FF-type model may be misspecified. To address such misspecifications, we propose modeling cross-correlations using a suitable structural model. Motivated by the recent clustering model (Nagy and Ormos 2018) and recursive model (Basak et al. 2018), we propose a social network dependence structure. Applied to data on stock returns for the 30 current DJIA stocks, we find evidence of network effects, the careful

---

1   Abnormal return is defined in the event study finance literature as the difference between the actual return of a security (in our case, over a one week time horizon) and the expected return as calculated using a model; see, for example, Brown and Warner (1985). Thus, any misspecification in the underlying factor model implies mismeasurement of expected returns and the corresponding risk-return relationship and would be evident in substantial abnormal returns.

modeling of which addresses misspecification of the underlying factor model. This brings returns more in line with risks, and provides a structural understanding of risk spillovers.

Any model is necessarily an abstraction of reality, and will entail a certain degree of misspecification; understandably this is true of the FF model as well. Researchers have continued to improve upon the FF model with a larger collection of factors (FF-type models), and this has undoubtedly improved model fit and interpretation. Our contribution here lies in proposing quite a different extension. We consider trading activity and its structural interpretations more explicitly than the literature, which goes along the lines of structural interpretation of correlations that currently lies beyond the scope of the FF-type models. Then, together with the factors in the FF model, the proposed model provides substantial enhancement to our understanding and a better explanation of returns. We consider as benchmark a CAPM model (including only the market return factor) and an FF-type model including 6 factors: the 5 Fama and French (1988, 1993) factors, plus the momentum factor of Carhart (1997). Our results show that the base CAPM, together with network effects, has competitive explanatory power, and for some stocks offers substantial improvements relative to the above 6 factor model. This provides an alternate structural factor model for asset pricing, and develops avenues for new research.

Section 2 contrasts factor models in finance and econometrics and draws some insights into misspecification. Section 3 develops a social network model and estimates this using the DJIA stock returns. Structural interpretation of the model is discussed, together with alternate structural models. Section 4 concludes, with an appeal for further research on structural dependence in stock returns.

2. Factor Models in Finance and Econometrics

The FF and similar factor models in finance are typically expressed as:

$$y_{it} = \alpha_i + \beta_i m_t + \gamma_i' x_t + \varepsilon_{it}, \tag{1}$$

and estimated using time series data ($t = 1, ..., T$) on the returns, y_{it}, on n ($i = 1, ..., n$) stocks. Here m_t denotes the excess return on the market portfolio and β_i the corresponding beta-factor for stock i, x_t is a vector of returns on a finite number of firm-specific factors (typically called the Fama-French factors) and γ_i their corresponding factor exposures, α_i is a stock- (firm-) specific intercept that can be interpreted as a fixed effect, and ε_{it} is an idiosyncratic error term. Typically, the excess return on the market portfolio (m_t) is easily computed from market data and the time-varying returns (x_t) are reported in market research publications (French 2017). The original Fama and French (1993) factors SMB (Small Minus Big) and HML (High Minus Low) were extended to include Mom (momentum) in Carhart (1997) and further to RMW (Robust Minus Weak) and CMA (Conservative Minus Aggressive) in Fama and French (2015). Returns on these factors constitute x_t; see French (2017) for further details on concepts and computation.

2.1. Network Effects and Bias

Traditionally, the above FF model (1) is estimated by least squares, where the factor exposures β_i and γ_i are viewed as parameters to be estimated from the data. This estimation strategy raises issues that are well recognized in the literature; see, for example, Strong (1992), Kothari and Warner (2007) and Kolari and Pynnönen (2010). One important issue is that risk is not consistently estimated if there is either time-varying volatility or cross-section correlations in the errors ε_{it}. This renders inference on abnormal returns particularly challenging. This is really an estimation efficiency issue that is not in itself likely to cause bias in estimation of the factor model. However, there would be a more serious problem of endogeneity if, for some reason, there were network interdependencies between returns on different stocks. This will also lead to mismeasured risk and very likely biased estimates.

To understand the nature of the problem, consider for simplicity a CAPM type restricted factor model of the form

$$y_{it} = \alpha_i + \beta_i m_t + \varepsilon_{it}, \tag{2}$$

where the effect of additional FF firm-specific factors is not included. The above CAPM model (2) can imply a specific form of network architecture, known in the spatial econometrics literature as a social interactions model (Lee et al. 2010; Hsieh and Lee 2016; Bhattacharjee et al. 2018; Cohen-Cole et al. 2018; Doğan et al. 2018) or a farmer-district model (Case 1992; Robinson 2003; Gupta and Robinson 2015), whereby the units (here, stocks) are classified into several groups or social networks. Stocks in the same social network are related to each other, but not to stocks in the other networks. Further, the inter-network influences are symmetric across all directed pairs of stocks within the network and can be represented by an adjacency-based binary weights matrix. In turn, the membership of social networks is inferred either by cluster analysis (Bhattacharjee et al. 2016; Chakraborty et al. 2018; Nagy and Ormos 2018) or correlation analysis (Junior et al. 2015; Bailey et al. 2016) of the dependent variable, which in our case are stock returns.

Now, consider the clustering pattern implied by the above CAPM, assuming for simplicity that the parameter vector can take only one of two values, $(\alpha_i, \beta_i) \in \{(-\frac{1}{2}, 1), (\frac{1}{2}, 0)\}$, and these correspond to the two network classes. Likewise, assume for simplicity only two time periods, $t = 1, 2$. Then, if a scatterplot of returns is obtained along the axes given by $t = 1$ and $t = 2$, it is clearly seen that the loci of data points in the two network classes will be $\left(m_1 - \frac{1}{2}, m_2 - \frac{1}{2}\right)$ and $\left(\frac{1}{2}, \frac{1}{2}\right)$, with the random distribution of points around the loci determined solely by the idiosyncratic errors ε_{it}.

More generally, if the parameter vector takes values in a finite set $(\alpha_i, \beta_i) \in \{(a_1, b_1), (a_2, b_2), \cdots, (a_k, b_k)\}$, and we have data for $t = 1, ..., T$ time periods, then this would generate data points clustered around a corresponding set of k loci: $\{(a_j + b_j m_1, a_j + b_j m_2, \cdots, a_j + b_j m_T) : j = 1, \cdots, k\}$. Further, if there were no network interdependence between the stocks, the parameters can be recovered through time series least squares regressions based on (2) for each individual stock, since the observations are independent over time. Therefore, the loci of the clusters can also be precisely estimated as the number of time periods increase, that is, as $T \to \infty$. The same argument holds if we had additional FF firm-specific factors, in which case we would estimate a model of the Fama-French form (1). However, the endogenous network effects would lead to biased least squares estimation.

Let us now consider a simple extension to the CAPM type restricted factor model (2) to include network interdependence. Denote by $\underline{y}_t = (y_{1t}, y_{2t}, \cdots, y_{nt})\prime$ the vector of returns at time point t, and consider a standard spatial autoregressive (lag) network model of the form

$$\underline{y}_t = \rho W \underline{y}_t + \underline{\alpha} + \underline{\beta} m_t + \underline{\varepsilon}_t, \tag{3}$$

where $\underline{\alpha}$ and $\underline{\beta}$ are corresponding vectors of the CAPM parameters, $W_{(n \times n)}$ is a square matrix of network membership with zero diagonal elements and the off-diagonal elements are unit if two firms belong to the same network and zero otherwise, and ρ is the so-called spatial autoregressive or network dependence parameter ($|\rho| < 1$). Here, the network architecture follows exactly the social interactions or farmer-district model; see, for example, Lee et al. (2010). Further, we assume as before that the CAPM parameter vector takes values in a finite set

$$(\alpha_i, \beta_i) \in \{(a_1, b_1), (a_2, b_2), \cdots, (a_k, b_k)\},$$

and further that two stocks with the same parameters belong to the same network.

Now, without loss of generality, let the stocks in the first network have parameters (a_1, b_1) and come first in the ordering, followed by the second network with parameter values (a_2, b_2) and so on, we have the following block-diagonal equicorrelation structure for W:

$$W = \begin{pmatrix} W_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & W_k \end{pmatrix},$$

where the connection matrix for network j $(j = 1, \ldots, k)$ is of order $(n_j \times n_j)$ and takes the form:

$$W_j = \begin{bmatrix} 0 & 1/(n_j - 1) & \cdots & 1/(n_j - 1) \\ 1/(n_j - 1) & 0 & \cdots & 1/(n_j - 1) \\ \vdots & \vdots & \ddots & \vdots \\ 1/(n_j - 1) & 1/(n_j - 1) & \cdots\cdots\cdots\cdots & 0 \end{bmatrix}_{(n_j \times n_j)},$$

with zero diagonal and row-standardised unit values everywhere else, and $n = n_1 + n_2 + \cdots + n_k$.

The equicorrelation form of the social interactions model clearly highlights why it may be useful to identify the network structure based on cross-section correlations, as in Lee et al. (2010) and mboxciteauthorB3-jrfm-450549 (2016). Then, the reduced form the network model (3) is:

$$\underline{y}_t = (I - \rho W)^{-1}\underline{\alpha} + (I - \rho W)^{-1}\underline{\beta}m_t + (I - \rho W)^{-1}\underline{\varepsilon}_t, \tag{4}$$

where

$$(I - \rho W)^{-1} = \begin{pmatrix} (I - \rho W_1)^{-1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & (I - \rho W_k)^{-1} \end{pmatrix}.$$

This reduced form representation (4) clearly highlights how the network generates risk spillovers.

The structure of the reduced form in (4) has important implications for estimation and inference on the Fama-French model. First, additional FF-type firm-specific factors retain the same basic structure of the model, with slope parameters that are proportional to the underlying structural model at (3). Second, applied to data from the network interactions model (3), time series least squares regression based on individual stock returns will simply recover the reduced form intercept and slopes, rather than the underlying structural parameters (a_j and b_j); clearly this leads to biased and inconsistent inferences. Third, the reduced form least squares parameter estimates correctly recover the underlying network structure because the nature of clustering does not change. Specifically, the data points are still clustered around a set of k loci, which is a simple scale transformation of the original model without network dependence. Hence, the network structure can be accurately identified by cluster analysis of the underlying returns. In fact, within the context of the FF model (1) with network dependence as in (3), cluster analysis will typically recover the loci of the FF- firm specific factors as well. Fourth, simply accounting for the network structure does not help. True, the underlying network dependence can be identified by clustering; but under the network interactions model (3), if the CAPM (and FF) part of the model were ignored, this would provide biased inferences on the network dependence parameters. Hence, both parts of the model are important for accurate estimation and analyses of risk and return.

2.2. Comparison with Panel Data Factor Model

Cross-section dependence is well studied within the current literature in panel data econometrics. Here, the central factor model has the following form:

$$y_{it} = \delta_i' f_t + \theta_i' z_{it} + \varepsilon_{it},\qquad(5)$$

where f_t denotes a vector of time-specific *"factors"* with corresponding stock-specific loadings δ_i, z_{it} contains a collection of stock- and time-varying covariates, and ε_{it} are stationary but potentially cross-section dependent and autocorrelated regression errors; see, for example, Pesaran (2006). Some of the "factors" may be observed, and others latent. In particular, a factor taking unit value in each time period corresponds to fixed effects, denoted α_i in the FF model (1). The response variable (returns) y_{it} has cross-section dependence arising from two sources. First, there is the influence of common "factors" f_t, but potentially with effects heterogeneous across different stocks. Second, there are the cross-section dependent errors ε_{it}.

Pesaran (2006) points to an important distinction between cross-section strong and weak dependence (of returns on different stocks, in our case). The first arises from the effect of common factors, such as the market portfolio and the FF factor returns; the second is due to local interdependencies (spillovers) between firms and their stocks. Following from Pesaran (2006), an influential literature has spawned in this area; see, for example, Kapetanios and Pesaran (2007), Bai (2009), Pesaran and Tosetti (2009), and Bailey et al. (2016).

Pesaran (2006) developed two key results. First, least squares estimation of (5) with omitted latent factors provides inconsistent and biased estimates of θ_i in general. The only situation where credible inferences can be made is when the errors ε_{it} are stationary over time and granular across the cross-section. Pesaran (2006) and Pesaran and Tosetti (2009) provide technical definitions of cross-sectional granularity. This is conceptually akin to stationarity, but across the cross-section dimension, and implies that the degree of cross-section dependence is limited. Pesaran (2006) terms this case weak cross-section dependence, and Pesaran (2015) provided a statistical test based on average cross-section correlation of the residuals; see Bhattacharjee and Holly (2013) for an alternate test.

Second, Pesaran (2006) offers a large sample method to address strong dependence when both dimensions are large, that is, $n \to \infty$ and $T \to \infty$. In such situations, one can enrich the model by including cross-section averages of the dependent and independent variables as:

$$y_{it} = \delta_i'(\bar{y}_t, \bar{z}_t) + \theta_i' z_{it} + \varepsilon_{it},\qquad(6)$$

where the cross-section averages $(\bar{y}_t, \bar{z}_t)$ eliminate strong dependence from the model, leaving only weak dependence in the residuals. This model can then be consistently estimated by least squares. This methodology is called common correlated effects estimation because $(\bar{y}_t, \bar{z}_t)$ take these high cross-section correlations out of the data.

Let us now revert to a comparison with the FF model (1). The return on the market is akin to the average return in each period, and hence is very close to $\bar{y}_t$. Since temporal variation in the risk free rate is much lower than the market, the excess return on the market, m_t, is numerically almost the same cross-section average return less a constant.[2] Unfortunately, beyond $\bar{y}_t$, common correlated effects cannot be directly applied in (1), because there are no regressors with both cross-section and time variation, unlike z_{it} in (5). This key observation has two implications. First, one should always test the residuals from least squares estimation of (1) for potential strong dependence. Second, strong cross-section dependence needs to be modelled based on structural considerations of pricing in financial markets. We focus on this second issue in the next section.

---

2   Over the period of our analysis, standard deviation of the risk free rate is only 0.15, as compared to 4.43 for the market return.

3. Structural Models of Asset Pricing Correlations

As discussed in Section 2, the lack of cross-section variation in the regressors precludes the opportunity to apply common correlated effects estimation in the FF model (1). This implies that any cross-section strong dependence (across the stocks) needs to be modelled explicitly using structural models of price formation in the market. This is where we turn to next. We first discuss a few structural models and the relevant literature, and propose an alternate model. Then, we illustrate our proposed structural model using data on monthly returns on the stocks currently included in the Dow Jones Industrial Average (DJIA).

3.1. Structural Models of Price Formation

Structural modelling of cross-section dependence is the domain of spatial and network econometrics. Within this literature, there is very little research on stock returns. However, one can draw some insights from the literature on other markets (for example, housing and labor markets) or to dependence across financial markets. A key result from this literature is that the underlying structural model is, in general, not fully identified from cross-section covariances and correlations (Bhattacharjee and Jensen-Butler 2013). Hence, one requires further structural assumptions, either from theory or from context specific research, to identify network effects.

One such admissible assumption is recursive structure under which information flows are sequential (but contemporaneous) through different segments of the market. Based on 25 portfolios formed on size and book-to-market (Fama and French 1996; French 2017), Basak et al. (2018) find substantial explanation for risk spillovers and abnormal returns, and the model outperforms reduced form VAR (vector autoregressive) factor models. Suppose risk neutral traders arrive sequentially and repeatedly at the market taking positions on preferred risk/return FF portfolios. Then, limit or market order mechanisms would generate such recursive ordering of the portfolios in terms of information flow, in turn leading to cross-portfolio correlations. This ties in with the recent market microstructure literature on limit orders; see, for example, Handa and Schwartz (1996), Parlour (1998), Foucault (1999), Mertens (2003) and Foucault et al. (2005).

Recursive ordering is often a feature of cross-market information flows, where the sequencing of opening and closing times of different markets produce the so-called 'meteor shower' phenomenon documented in volatility by Engle et al. (1990) and in returns by Hamao et al. (1990). Bhattacharjee (2017) find that return correlations across 19 markets worldwide are explained by recursive ordering, in combination with global factors capturing the dominance of major markets. We propose a different model in this paper, but recursive structural models hold good promise for the future.

Beyond recursive ordering, some other structural assumptions are also admissible. Bhattacharjee and Jensen-Butler (2013) show that network dependence structure is identified under the assumption of symmetric interdependence. There are three important special cases of symmetry. First, there are social network models where individuals in the same network share information or interact with each other, but not with members in other networks; see, for example, Lee et al. (2010) and Cohen-Cole et al. (2018). This model is also closely related to the farmer-district model (Case 1992; Robinson 2003). Cross-section dependent stock returns can well be represented by social network models, but in applications, the membership and identity of networks is seldom known *a priori*, and one still needs appropriate theory or clustering/LASSO methods (discussed below) to motivate these networks.

Second, there are models where interconnections are binary and reciprocal, as in the social networks model, but the networks can be overlapping. In the context where the network is sparse, but negative interactions are possible, Bailey et al. (2016) propose estimation of network structure based on multiple testing of estimated cross-section correlations. We will briefly consider a model of this type later. In the context of stock returns factor models, lack of any obvious structural interpretation of the network is obviously an impediment. Besides, one would expect network interdependence in the stock market to be fairly dense, and hence the sparsity assumption may also be somewhat tenuous.

The third class of models use either clustering or LASSO (least absolute shrinkage and selection operator; Tibshirani 1996) to identify social network or block structure from the data. However, an observed pattern of clustering does not necessarily imply clustering of the parameter vector, since the returns of different stocks may be similar because they share a network. This observation justifies current approaches of characterizing the underlying latent network structure using either spatial and spatio-temporal clustering (Bhattacharjee et al. 2016; Chakraborty et al. 2018) or an analysis of cross-section correlations (Junior et al. 2015; Bailey et al. 2016); Nagy and Ormos (2018) apply clustering across markets to study dependence. Lam and Souza (2016) provide a method to identify block structure using the LASSO and similar methods. Our structural model and empirical analysis are based mainly on this third approach. As discussed earlier, since the network structure can only contribute to weak dependence, the factor structure needs to be removed from the data a priori. As our analytical discussion in Section 2.1 shows, one can first estimate a FF-type factor model, identify the common factors that are associated with strong dependence, and then cluster the stocks based on the remaining weak-dependence factors. However, since the factors in the FF model are essentially returns on different types of risk, clustering can then be based on exposure to the corresponding risks.

The above approach is consistent with the following trading strategy. In the context of factor models (1) and (5), absence of cross-section variation in regressors for the FF-type model imply that subtle variations, over time, in the factor exposures for each stock are not captured in the data; hence, traders need to address this issue through diversification. Then, our structural model posits that, traders sort themselves on heterogenous risk preferences. Given a specific preference type over multiple risk factors, they then choose their preferred exposure to the risks and create a diversified portfolio of stocks with this risk exposure. This behavior generates interdependence across stock returns within this portfolio, but not beyond. Since the exposures are estimated by an FF-type factor model (1), the network can be identified by clustering the stocks on this estimated exposure vector.[3] The above trading model is structural, but its assumptions need validation. To highlight the promise that this approach holds, we now provide an illustrative application on the DJIA stock returns.

Our central argument is that, since the structural network effects are only partially identified from reduced form regressions (Bhattacharjee and Jensen-Butler 2013), inference requires structural assumptions underpinned by appropriate theory. Above, we have discussed three such lines of theory, emphasizing in particular one new structural model. There, traders choose their diversified portfolio with a preferred risk exposure; this trading behavior generates interdependence across stock returns within this portfolio, but not beyond. Since the exposures are estimated by an FF-type factor model, the network can be identified by clustering the stocks on this estimated exposure vector. This justifies our approach (in Section 3.2) of assuming that within cluster correlations dominate network effects across the clusters, which are then assumed to be absent.

Obviously, there are competing structural models where stocks belonging to different groups would be correlated, and we also discussed two such models: First, we refer to Basak et al. (2018), who developed a model where limit or market order mechanisms generate recursive ordering of the portfolios in terms of information flow, in turn leading to cross-portfolio correlations. This may be viewed as a model diametrically in opposition to the model developed in this paper. Second, unrestricted correlations can be modeled, and we also consider this approach. However, the unrestricted correlations model raises two further issues from a structural point of view. First, we need an assumption of sparsity (Bailey et al. 2016). Network interdependence in the stock market may be dense, and hence the sparsity assumption may be somewhat tenuous. Second, we do not currently have theory to justify sparse interactions, and lack of any obvious structural interpretation of the network is obviously an impediment. Nevertheless, we estimated such a model (Section 3.2)

---

[3] Alternate structural restrictions with asymmetric dependence, for example, tree-based nested dependence (Bhattacharjee and Holly 2013), sparsity (Ahrens and Bhattacharjee 2015; Lam and Souza 2018) and copulas (Liu et al. 2018) can also hold promise, but we do not consider these here.

and highlight that more work is required for structural understanding of the underlying trading mechanisms. We suggest this as an avenue for future research.

3.2. Data and Estimated Model

We collected daily stock returns data (adjusted for splits, dividends and distributions) from Yahoo Finance, on the stocks currently included in the Dow Jones Industrial Average (DJIA).[4] The period under analysis is January 2001 to December 2015.[5] Historical monthly factor returns on the Fama and French (1993, 2015) 5-factors and the Carhart (1997) momentum factor were collected from the web archive of French (2017). To make our stock returns data comparable with factor returns, stock returns are aggregated to the monthly level. These constitute our data under analysis.

First, we estimate by least squares a CAPM model (2) including only an intercept and excess return on the market. As discussed in Section 2.1, the network structure can be accurately identified by clustering on the vector (α_i, β_i). We also estimate a FF-type factor model including all the six factors: m_t, SMB, HML, RMW, CMA and Mom. The CAPM model exhibits spatial (network) strong dependence. Using the CD test of Pesaran (2015), the null hypothesis of weak dependence is strongly rejected. The test statistic evaluates to -4.026 with a *p*-value of 5.7×10^{-5}. However, the same is not true for the FF-type model with six factors; the CD test statistic is -1.775 with a *p*-value of 0.076.

Next, we apply cluster analysis to the estimated $(\hat{\alpha}_i, \hat{\beta}_i)$, and clearly identify 3 clusters: low alpha and low beta (13 stocks), low alpha and high beta (13 stocks) and high alpha (4 stocks). The membership of the clusters is reported in Table 1. Then, we construct a social network weighting matrix $W_{(n \times n)}$ based on membership of the above three clusters.

Finally, we estimate two spatial autoregressive (lag) network models, including only m_t but not the Fama and French (1993, 2015) or Carhart (1997) factors. In the first, a contemporaneous spatial lag $W\underline{y}_t$ is included; this is exactly the model in (3). Inclusion of the spatial lag introduces endogeneity in the model, and we estimate using a variant of the popular two stage least squares (2SLS) method in Kelejian and Prucha (1998). Like the application of common correlated effects, the above 2SLS method also presents challenges because there is no cross-section variation in the regressors. We use as instruments the omitted five FF-type factors, together with lagged residuals from the estimated CAPM model (2). The first stage estimation work well (with F-statistics much greater than 10 in all cases) and weak instruments issues are not apparent. However, 2SLS is known to have finite sample bias and there is loss of efficiency; for this reason, model comparison is based on root mean squared errors (RMSE). The second network model includes network effects as a time-lag, that is $W\underline{y}_{t-1}$. Here, we do not have contemporaneous endogeneity and the model can be estimated using least squares:

$$\underline{y}_t = \rho W \underline{y}_{t-1} + \underline{\alpha} + \underline{\beta} m_t + \underline{\varepsilon}_t, \tag{7}$$

The spatial lag (3) and space-time lag (7) model would in general have different structural implications. However, in our specific context, they are similar since the time lag is one week, which is very long in financial markets. By this time lag, all stock specific temporal information is expected to already have been factored into prices, and this information is therefore not relevant for trading strategy based on portfolio construction. Under our proposed structural model discussed in Section 3.1, trading behavior generates network effects through the choice of portfolios which get updated at a much lower frequency. Hence, structural implications of the spatial autoregressive and time lag models

---

4   The following 30 stocks are included (tickers in parentheses): 3M (MMM), American Express (AXP), Apple (AAPL), Boeing (BA), Caterpillar (CAT), Chevron (CVX), Cisco Systems (CSCO), Coca-Cola (KO), DowDuPont (DOW), ExxonMobil (XOM), Goldman Sachs (GS), The Home Depot (HD), IBM (IBM), Intel (INTC), Johnson & Johnson (JNJ), JPMorgan Chase (JPM), McDonald's (MCD), Merck & Company (MRK), Microsoft (MSFT), Nike (NKE), Pfizer (PFE), Procter & Gamble (PG), Travelers (TRV), UnitedHealth Group (UNH), United Technologies (UTX), Verizon (VZ), Visa (V), Walmart (WMT), Walgreens Boots Alliance (WBA), and Walt Disney (DIS).

5   Data for Visa (V) are only from March 2008. Our methods are applicable to unbalanced panel data.

are literally the same in so far as network effect implications are concerned. In terms of econometric implications, estimation of the two models are different. The spatial lag model generates endogenous effects; hence, we use instrumental variables methods, while the space-time lag model has only lagged endogenous effects, and therefore, least squares estimation is employed.

Table 1. Clusters, Model Diagnostics and Relative Efficiencies.

Clusters (Network)	Ticker	Root Mean Squared Errors (RMSE)				Efficiency, Relative to	
		CAPM (2)	FF Model (1)	Contemporaneous Network (3)	Time-Lag Network (7)	CAPM (2)	FF Model (1)
Low alpha, low beta							
	PFE	4.757	4.596	**4.357**	4.744	−8.41%	**−5.20%**
	TRV	4.782	4.752	**4.633**	4.786	−3.11%	**−2.50%**
	MCD	4.764	4.700	**4.630**	4.777	−2.81%	**−1.50%**
	JNJ	3.697	3.445	**3.437**	3.660	−7.02%	**−0.23%**
	XOM	4.380	4.226	**4.224**	4.384	−3.55%	**−0.03%**
	KO	4.308	4.058	**4.066**	4.319	−5.61%	0.21%
	WMT	4.827	4.759	4.966	**4.819**	−0.17%	1.26%
	PG	3.927	3.720	**3.772**	3.932	−3.94%	1.39%
	WBA	6.515	6.325	6.788	**6.496**	−0.30%	2.71%
	MMM	4.431	4.263	4.524	**4.383**	−1.08%	2.81%
	CVX	4.807	4.603	**4.733**	4.798	−1.54%	2.84%
	MRK	6.489	5.939	**6.329**	6.501	−2.46%	6.56%
	VZ	5.536	4.969	5.570	**5.479**	−1.03%	10.27%
High alpha							
	V	6.217	5.894	5.490	**5.432**	−12.64%	**−7.84%**
	UNH	6.693	6.699	**6.658**	6.704	−0.52%	**−0.61%**
	NKE	5.681	5.657	6.508	**5.688**	**0.14%**	0.56%
	AAPL	9.143	8.929	9.906	**9.112**	−0.34%	2.05%
Low alpha, high beta							
	BA	6.340	6.262	6.316	**6.065**	−4.33%	**−3.15%**
	UTX	4.403	4.348	**4.322**	4.402	−1.82%	**−0.58%**
	DIS	4.760	4.688	4.708	**4.666**	−1.98%	**−0.47%**
	HD	5.680	5.698	5.767	**5.683**	0.05%	**−0.28%**
	MSFT	5.730	5.651	5.696	**5.733**	−0.58%	0.81%
	GS	5.938	5.856	5.946	**5.937**	−0.03%	1.38%
	INTC	7.283	6.887	7.224	**7.202**	−1.11%	4.56%
	CAT	6.501	6.201	6.518	**6.498**	−0.04%	4.78%
	IBM	5.303	4.937	5.322	**5.175**	−2.41%	4.83%
	CSCO	7.561	6.946	**7.395**	7.438	−2.19%	6.47%
	DOW	8.606	7.382	8.641	**8.540**	−0.76%	15.69%
	AXP	7.349	6.200	7.320	**7.255**	−1.27%	17.03%
	JPM	6.401	5.370	6.852	**6.418**	**0.26%**	19.52%

Between the two network models (3) and (7), we choose the one with lower RMSE; the model with better fit is indicated in bold in Table 1. Then, we apply the CD test to the correlation matrix of residuals. The test statistic is 1.458 with a *p*-value of 0.145. Hence, we are satisfied that weak dependence holds, and estimates of the network factor model are consistent. Finally, we report relative efficiency of the chosen network model, in terms of percent lower RMSE, relative to the CAPM model (1) and the FF-type model (2).

In terms of RMSE, the clustering network model improves upon the CAPM for all stocks except two (Nike and JPMorgan Chase). This is reassuring but not surprising because the network model includes one addition regressor, the spatial lag. However, it is remarkable that the network model improves upon the FF-type model with all six factors for 11 (out of the 30) stocks. This provides encouraging validation of the clustering structural model proposed in this paper. Understanding trading activity and pricing in financial markets is an important problem in finance. It is our belief that the work here takes an important step in this direction.

In Table 2, we report the estimated alphas (α) and betas (β) for the CAPM and network models. The distinction between the three estimated clusters is clear from the CAPM estimates. Further, as predicted by theory, there is strong correlation between the estimates from the two models; 0.68 for alpha and 0.76 for beta. However, also as expected from our theory, there is substantial bias in the CAPM model estimates; on average, positive bias in alpha is about 66% and 30% for beta. The FF-type model with 6 factors is qualitatively similar. Since the time-period under study is not too long, we assume that W and ρ is constant over time, but that ρ varies by stock (that is, ρ_i).

Table 2. Factor Model Estimates.

Clusters (Network)	Ticker	Estimated Network Model—(3) or (7)			Estimated CAPM	
		Alpha	Beta	Network (rho)	Alpha	Beta
Low alpha, low beta						
	PFE	−0.061	0.669	−0.147	−0.162	0.650
	TRV	0.463	0.787	−0.090	0.369	0.771
	MCD	0.746	0.660	0.013	0.670	0.649
	JNJ	0.515	0.484	−0.172	0.374	0.457
	XOM	0.329	0.582	−0.078	0.249	0.567
	KO	0.231	0.524	−0.032	0.179	0.515
	WMT	−0.224	−0.115	0.786	0.043	0.388
	PG	0.493	0.395	−0.064	0.409	0.380
	WBA	−0.033	0.044	1.222	0.305	0.779
	MMM	0.299	0.415	0.668	0.427	0.808
	CVX	0.247	0.298	0.703	0.410	0.721
	MRK	0.008	0.615	−0.080	−0.107	0.596
	VZ	0.254	0.696	−0.264	0.162	0.676
High alpha						
	V	0.698	0.414	0.355	1.410	0.785
	UNH	1.023	0.607	0.049	1.090	0.614
	NKE	0.060	0.223	0.608	1.133	0.762
	AAPL	0.716	0.248	1.423	2.612	1.252
Low alpha, high beta						
	BA	0.225	1.045	0.307	0.391	1.090
	UTX	0.311	0.976	0.055	0.329	0.981
	DIS	0.446	1.014	0.122	0.488	1.174
	HD	0.375	0.465	0.428	0.453	1.020
	MSFT	0.282	1.042	−0.062	0.430	1.059
	GS	−0.063	0.943	0.353	0.053	1.396
	INTC	−0.124	1.184	0.163	−0.015	1.402
	CAT	0.429	0.847	0.520	0.461	1.494
	IBM	−0.119	0.486	0.341	0.100	0.953
	CSCO	−0.173	1.599	−0.246	−0.319	1.563
	DOW	0.168	1.629	0.210	0.277	1.656
	AXP	−0.075	1.442	0.215	0.008	1.463
	JPM	0.025	1.367	0.025	0.152	1.388

In addition to the clustering model, we also applied the multiple testing procedure of Bailey et al. (2016) to construct a weighting matrix; in this case, the *p*-value of the CD test is 0.650, which is promising performance of this method in negating strong dependence. However, more work is required for structural understanding of the underlying trading mechanisms; this is also an avenue for future research.

We verify the robustness of our findings across several dimensions. First, we evaluate robustness in clustering. We use different starting clusters, and different algorithms, all of which provide consistent results. Further, we account for the uncertainty in estimated alpha and beta parameters by multiple imputation based on estimated confidence intervals, and the results are consistent as well. Second, we evaluate robustness in choice of factors by considering two traditional factor models. One is the CAPM with a single market return factor, and the other is a 6-factor model including the Fama and French (1988, 1993) and Carhart (1997) factors. We find that our model almost always provides an improvement over the CAPM (in terms of RMSE), but is also frequently better than the 6-factor model. Third, we consider several network models. One, a contemporaneous spatial lag model; two, a space-time lag model; and finally, a sparse network model with unrestricted interactions. The main implications of our results are consistent across all three specifications.

4. Conclusions

The Fama and French (1993) and similar factor models are important and popular in finance, and they provide good structural understanding of the risk, returns and price formation. Typically, the model is estimated as a time series regression separately for each stock (firm). Such estimation would provide consistent estimates if the data are independent across firms. However, if there were any network effects, such estimates can be inefficient or even inconsistent if the network effects are endogenous.

Indeed, persistent evidence of abnormal returns and cross-section correlations in stock returns points towards potential misspecification of the FF-type models. In this paper, we show that endogenous network effects create cross-section dependence that renders least squares estimation of FF-type factor models inconsistent; hence, computed returns and risk may both be erroneous.

Further, we argue that current econometric methods to deal with cross-section dependence are not applicable to the above factor models. This leads us to development of structural models to understand network effects better. We propose a social network model based on clustering and show that it lends itself to interesting structural interpretations. Applied to data on the 30 DJIA stocks, our model provides improved estimation of factor models and insightful new understanding of trading activity and price formation. How the information in improved relative efficiencies can be harnessed for trading is a matter of further research and practice, which we also retain for future work.

While our current evidence is limited to only the DJIA stocks, this work provides the basis for further empirical validation and development of theory, not to mention alternate structural models of trading activity as well. A larger temporal dimension would obviously be useful in highlighting the weaknesses of the FF model which ignores structural cross-sectional interactions that are highlighted from our findings. However, capturing such interactions requires a potentially strong assumption that the nature and strength of interactions is constant over time. Obviously, the validity of this assumption would become more tenuous with a larger sample, but can equally be verified using more data. The advantages of larger sample data would also be apparent with a larger cross-section dimension. The current paper is best viewed as a proof of concept that further research on structural network effects may be fruitful. Hence, our work provides several promising avenues for further research in the direction of market microstructure models and their applications.

References

Ahrens, Achim, and Arnab Bhattacharjee. 2015. Two-step Lasso estimation of the spatial weights matrix. *Econometrics* 3: 128–55. [CrossRef]

Bai, Jushan. 2009. Panel data models with interactive fixed effects. *Econometrica* 77: 1229–79.

Bailey, Natalia, Sean Holly, and M. Hashem Pesaran. 2016. A Two-Stage Approach to Spatio-Temporal Analysis with Strong and Weak Cross-Sectional Dependence. *Journal of Applied Econometrics* 31: 249–80. [CrossRef]

Basak, Gopal K., Arnab Bhattacharjee, and Samarjit Das. 2018. Causal ordering and inference on acyclic networks. *Empirical Economics* 55: 213–32. [CrossRef]

Bhattacharjee, Arnab. 2017. Identifying the Causal Structure of Directed Acyclic Graphs (DAGs). Paper presented at the 40th International Panel Data Conference, Thessaloniki, Greece, July 7–8.

Bhattacharjee, Arnab, and Sean Holly. 2013. Understanding interactions in social networks and committees. *Spatial Economic Analysis* 8: 23–53. [CrossRef]

Bhattacharjee, Arnab, and Chris N. Jensen-Butler. 2013. Estimation of the spatial weights matrix under structural constraints. *Regional Science and Urban Economics* 43: 617–34. [CrossRef]

Bhattacharjee, Arnab, Eduardo A. Castro, Taps Maiti, and Zhen Zhang. 2016. Spatio-Temporal Patterns in Portuguese Regional Fertility Rates: A Bayesian Approach for Spatial Clustering of Curves. Mimeo. Available online: https://pureapps2.hw.ac.uk/ws/portalfiles/portal/10595008/POR_Demography_paper.pdf (accessed on 24 August 2018).

Bhattacharjee, Arnab, Sean Holly, and Jesus Mur. 2018. Contemporary developments in the theory and practice of spatial econometrics. *Spatial Economic Analysis* 13: 139–47. [CrossRef]

Boehmer, Ekkehart, Jim Masumeci, and Annette B. Poulsen. 1991. Event-study methodology under conditions of event-induced variance. *Journal of Financial Economics* 30: 253–72. [CrossRef]

Brown, Stephen J., and Jerold B. Warner. 1985. Using daily stock returns: The case of event studies. *Journal of Financial Economics* 14: 3–31. [CrossRef]

Carhart, Mark M. 1997. On Persistence in Mutual Fund Performance. *Journal of Finance* 52: 57–82. [CrossRef]

Case, Anne. 1992. Neighborhood influence and technological change. *Regional Science and Urban Economics* 22: 491–508. [CrossRef]

Chakraborty, Sayan, Arnab Bhattacharjee, and Taps Maiti. 2018. Structural Factorization of Latent Adjacency Matrix, with an application to Auto Industry Networks. Mimeo. Available online: https://www.researchgate.net/publication/324599853_Latent_Space_Linkages_of_Three_US_Auto_Manufacturing_Giants (accessed on 24 August 2018).

Chiang, Thomas C., and Jiandong Li. 2012. Stock returns and risk: Evidence from quantile. *Journal of Risk and Financial Management* 5: 20–58. [CrossRef]

Cohen-Cole, Ethan, Xiaodong Liu, and Yves Zenou. 2018. Multivariate choices and identification of social interactions. *Journal of Applied Econometrics* 33: 165–78. [CrossRef]

Doğan, Osman, Süleyman Taşpınar, and Anil K. Bera. 2018. Simple tests for social interaction models with network structures. *Spatial Economic Analysis* 13: 212–46. [CrossRef]

Engle, Robert F., Takatoshi Ito, and Wen-Ling Lin. 1990. Meteor Showers or Heat Waves? Heteroskedastic Intra-Daily Volatility in the Foreign Exchange Market. *Econometrica* 58: 525–42. [CrossRef]

Fama, Eugene F., and Kenneth R. French. 1988. Permanent and temporary components of stock prices. *Journal of Political Economy* 96: 246–73. [CrossRef]

Fama, Eugene F., and Kenneth R. French. 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33: 3–56. [CrossRef]

Fama, Eugene F., and Kenneth R. French. 1996. Multifactor explanations of asset pricing anomalies. *The Journal of Finance* 51: 55–84. [CrossRef]

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

Foucault, Thierry. 1999. Order flow composition and trading costs in a dynamic limit order market. *Journal of Financial Markets* 2: 99–134. [CrossRef]

Foucault, Thierry, Ohad Kadan, and Eugene Kandel. 2005. Limit order book as a market for liquidity. *The Review of Financial Studies* 18: 1171–217. [CrossRef]

French, K. R. 2017. Current Research Returns. Available online: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html (accessed on 24 August 2018).

Gupta, Abhimanyu, and Peter M. Robinson. 2015. Inference on higher-order spatial autoregressive models with increasingly many parameters. *Journal of Econometrics* 186: 19–31. [CrossRef]

Hamao, Yasushi, Ronald W. Masulis, and Victor Ng. 1990. Correlations in price changes and volatility across international stock markets. *The Review of Financial Studies* 3: 281–307. [CrossRef]

Handa, Puneet, and Robert A. Schwartz. 1996. Limit order trading. *The Journal of Finance* 51: 1835–61. [CrossRef]

Hsieh, Chih-Sheng, and Lung-Fei Lee. 2016. A social interactions model with endogenous friendship formation and selectivity. *Journal of Applied Econometrics* 31: 301–19. [CrossRef]

Junior, Leonidas Sandoval, Asher Mullokandov, and Dror Y. Kenett. 2015. Dependency relations among international stock market indices. *Journal of Risk and Financial Management* 8: 227–65. [CrossRef]

Kapetanios, George, and M. Hashem Pesaran. 2007. Alternative approaches to estimation and inference in large multifactor panels: Small sample results with an application to modelling of asset returns. In *The Refinement of Econometric Estimation and Test Procedures: Finite Sample and Asymptotic Analysis*. Edited by Garry D. A. Phillips and Elias Tzavalis. Cambridge: Cambridge University Press, pp. 239–81.

Kelejian, Harry H., and Ingmar R. Prucha. 1998. A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *The Journal of Real Estate Finance and Economics* 17: 99–121. [CrossRef]

Kolari, James W., and Seppo Pynnönen. 2010. Event study testing with cross-sectional correlation of abnormal returns. *Review of Financial Studies* 23: 3996–4025. [CrossRef]

Kothari, Sagar P., and Jerold B. Warner. 2007. Econometrics of Event Studies. In *Handbook of Empirical Corporate Finance*. Edited by B. Espen Eckbo. Amsterdam: Elsevier/North-Holland, pp. 3–36.

Lam, Clifford, and Pedro C. L. Souza. 2016. Detection and estimation of block structure in spatial weight matrix. *Econometric Reviews* 35: 1347–76. [CrossRef]

Lam, Clifford, and Pedro C. L. Souza. 2018. Estimation and Selection of Spatial Weight Matrix in a Spatial Lag Model. *Journal of Business and Economic Statistics*. forthcoming.

Lee, Lung-Fei, Xiaodong Liu, and Xu Lin. 2010. Specification and estimation of social interaction models with network structures. *The Econometrics Journal* 13: 145–76. [CrossRef]

Liu, Guizhou, Xiao-Jing Cai, and Shigeyuki Hamori. 2018. Modeling the Dependence Structure of Share Prices among Three Chinese City Banks. *Journal of Risk and Financial Management* 11: 57. [CrossRef]

Marks, Joseph M., and Jim Musumeci. 2017. Misspecification in event studies. *Journal of Corporate Finance* 45: 333–41. [CrossRef]

Mertens, Jean F. 2003. The limit-price mechanism: In honour of Martin Shubik. *Journal of Mathematical Economics* 39: 433–528. [CrossRef]

Nagy, László, and Mihály Ormos. 2018. Friendship of Stock Market Indices: A Cluster-Based Investigation of Stock Markets. *Journal of Risk and Financial Management* 11: 88. [CrossRef]

Parlour, Christine A. 1998. Price dynamics in limit order markets. *The Review of Financial Studies* 11: 789–816. [CrossRef]

Pesaran, M. Hashem. 2006. Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica* 74: 967–1012. [CrossRef]

Pesaran, M. Hashem. 2015. Testing weak cross-sectional dependence in large panels. *Econometric Reviews* 34: 1089–117. [CrossRef]

Pesaran, M. Hashem, and Elisa Tosetti. 2009. Large panels with spatial correlations and common factors. *Journal of Econometrics* 161: 182–202. [CrossRef]

Robinson, Peter M. 2003. Denis Sargan: Some perspectives. *Econometric Theory* 19: 481–94. [CrossRef]

Strong, Norman. 1992. Modelling abnormal returns: A review article. *Journal of Business Finance & Accounting* 19: 533–53.

Tibshirani, Robert. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B (Methodological)* 58: 267–88. [CrossRef]