Journal of Real Estate Finance and Economics, 29:1, 99–118, 2004

Dynamics of Urban Residential Property Prices—A Case Study of the Manhattan Market

ZHENG WANG
Deutsche Bank, New York, New York 10005, U.S.A.
E-mail: zheng.wang@db.com

Abstract

In a multivariate vector autoregression framework, this paper investigates the weak efficiency of the urban residential real estate market and the cause of weak efficiency. An error correction model is used to estimate long-term relationships among apartment prices and adjustment speed from disequilibrium to equilibrium. Based on a unique dataset of the Manhattan market, the efficiency of this market, seasonal stationarity of property prices, and the weak exogeneity of leading sub-markets are studied. Our results indicate that, in a market of less heterogeneity and higher transaction volumes, the weak efficiency hypothesis is rejected as in previous studies. This result implies that heterogeneity and lack of transaction information may not be the direct source of market inefficiency. Meanwhile, it is found that there are stable long-term relations among prices of different sub-markets. Interestingly, the price of one-bedroom co-operatives (co-op) is weakly exogenous. This implies that the starting co-op for most home buyers in urban areas is a leading indicator of the entire market, which contradicts the claim that high-end luxury co-op leads the market.

Key Words: cointegration, error correction model, equilibrium, exogeneity, heterogeneity, likelihood function, liquidity, market efficiency, seasonality, stationarity, vector autoregression

1. Background and motivation

The dynamics of house pricing has been of constant interest in real estate literature. Significant effort has been devoted to testing the efficiency of the housing market. Fama (1970) defined three types of efficiency based on the information content. What is relevant to this article is the weak efficiency. In a weakly efficient market, the excess return is zero when past prices and returns are available. Previous research, in general, rejected the weakly efficient hypothesis. Both house prices and excess returns exhibit serial correlation, which makes it possible to predict future prices and excess returns using historical prices and returns. Illiquidity of the market, which is related to high transaction cost; the heterogeneity of the apartments and difficulty in obtaining transaction information; shortage of supply due to high construction cost of housing; and uniqueness of being both consumption goods and investment vehicle, as well as government regulation were often cited to explain the inefficiency. Cho (1996) presented a comprehensive survey on these results.

Recent research has focussed on explaining the cause of inefficiency. In an empirical study, Abraham and Hendershott (1996) documented that prices in three northeast cities and eleven west-coast cities behaved differently from those in the rest of country. In a recent paper, Capozza et al. (2002) studied the determinant factors for housing price changes. The roles of information dissemination, supply constraints, and backward-looking expectation formation about market dynamics were investigated. In this paper, we also study the factors that have an impact on the weak efficiency. Instead of directly modeling those factors, we choose a special market for our study. This market lacks some of the attributes that were cited by researchers in explaining market inefficiency. The logic behind our methodology is that, if the weakly efficient hypothesis is rejected in this market, then attributes that have been cited in previous studies but do not present in this market may not be as important as those presented in this market.

We chose a very small market, the Manhattan residential real estate market, for our analysis. Analysis of this market provides some methodological advantages. First, all the transactions take place in a very limited geographic location. Although the location still plays an important role in determining apartment price, it is not as important as that in a wide-open geographic region. Second, the layout of apartments with the same number of bedrooms is uniformly similar. Therefore, they can be categorized into sub-markets by the number of bedrooms. Apartments in each sub-market have similar structure, size, and accessories. These two features reduce the heterogeneity of the properties to some extent, which is considered as one of the causes for market inefficiency. Third, the transactions in each sub-market are relatively active, which provides continuous observations for apartment prices. More important, because most of the transactions take place in apartment buildings, where the same type of apartments are traded from time to time, the price of one apartment can be easily implied from just-completed transactions. Therefore, we believe that transaction information is efficiently disseminated. Fourth, the transaction cost is relatively lower compared with that in other regions in the country. The seller usually pays the broker 6 percent commission and the buyer does not pay. In some cases, brokers provide further concessions to sellers when the dollar amount of the transactions is large. In a sense, the commission may be lower than a brokerage fee charged in some bond tradings. Therefore, the inefficiency cannot be directly related to commission. Overall, the Manhattan market provides a good sample for us to test whether market lack-of-transaction information, low explicit transaction cost, and property heterogeneity are the causes for market inefficiency.

The data analysis techniques used in this paper have two advantage over previous studies. First, we use a multivariate vector autoregression model (MVAR) to take advantage of our data in testing market efficiency. The price of each sub-market is modeled by past prices from all sub-markets. In this way, the test is more sensitive, such that weak efficiency can be rejected even if the price of one sub-market relies only on the prices of other sub-markets but not on its own history. Previous research focused on a single price index for the entire market and might have overlooked this possibility. Second, we shed light on the structural equilibrium among different sub-markets using the Error Correction Model (ECM) proposed by Engle and Granger (1987). The ECM allows us to identify long-term relationships among prices from different sub-markets and to isolate some sub-markets that lead other sub-markets, as well as to estimate how each sub-market adjusts itself back to equilibrium when a disequilibrium occurs.

Although most of the housing pricing models are time series models in the literature, the common practice is to construct a single price index as a proxy price for the market. Few researchers explicitly use multivariate time series models. Among those exceptions, Abraham and Hendershott (1996) used a simultaneous equation system to estimate housing prices and price changes. More related to our research, Malpezzi (1999) explored the relationship between housing prices and income, and Zhou (1997) studied housing prices and income in a simultaneous equation system. Both used an ECM of two equations. The common feature of ECM is that the long-term relationship and short-term adjustment are explicitly modeled. The long-term relationship is described as cointegration among related variables. The short-term adjustment includes the momentum from previous value and the correction of disequilibrium from long-term equilibrium.

Although both Malpezzi (1999) and Zhou (1997) used stationarity, unit roots, and cointegration in their studies, our model specification, modeling process, and result explanation are significantly different from theirs. First, we simultaneously model prices of five sub-markets in one model. Because of this, the cointegration among these prices can be more involved compared with a model with two endogenous variables. In a two-equation model, the possible cointegration is either one or zero. A cointegration can be expressed as a linear equation of two variables. In our model, the number of cointegrations can be as many as four. Each of them can be a linear equation with up to five variables. Therefore, identifying cointegration and the variables in each cointegration equation becomes an important part of our modeling process. Second, the model specification may vary depending on our estimates from the data. In an MVAR, it is possible that some variables may be weakly exogenous. The identification of weakly exogenous variables is part of cointegration identification. The existence of weak exogeneity changes the specification of the model, and therefore changes estimation and forecasting. Third, the explanation of long-term relationship may be more subtle. For a model of two equations, as in both Malpezzi (1999) and Zhou (1997), the two variables have clear economic meanings, and it is intuitive to explain their relationship. In our model, we implicitly consider an economy of five types of apartment. We do not explicitly model the consumer preference and producer technology of the economy. What we observe in our model are the equilibrium prices and transaction quantities, as well as the physical characters of the apartments. Because of this, we have to narrow our scope to focus on the relationships among prices of different sub-markets. We mainly rely on statistical testings to identify these relationships. To be consistent through the entire model, each relationship has to be tested in the presence of other identified relationships. The estimation is performed after all relationships are identified. Last, we study seasonal stationarity in the prices. A time series that exhibits seasonal non-stationarity has different behavior at different seasons. The presence of seasonal non-stationarity is another important price fluctuation source.

Using the median co-operatives (co-op) apartment prices of five sub-markets, we reject the hypothesis of market efficiency by showing that the price of each sub-market is closed related by its own lagged term and the lagged terms of other sub-markets. We identify two long-term relationships among these five sub-markets. Surprisingly, we find that the one bedroom sub-market plays a pivotal role in directing the overall market. We also find that the adjustment for the smaller co-ops market is much faster than that in a larger co-op market. Overall, the adjustment speed of the Manhattan co-op market is faster than the housing market reported by Meese and Wallace (1993).

In Section 2, we introduce the market and data. We present our model and (seasonal) stationarity test of the data in Section 3. In Section 4, we describe the model selection, cointegration identification, and model estimation processes. We also describe the identification of the weakly exogenous variable. The analysis in both Section 3 and Section 4 is based on median price. We perform similar analysis using average price and price per room. The results are similar to those based on median price. Results based on average price and price per room, and the comparison of these results based on three different price measures, can be obtained from the author. Section 5 concludes our study and further discusses the implications of our findings.

2. The Manhattan market and data

The layouts of Manhattan apartments are more or less uniformly standard. For example, a studio apartment usually has a living room, which also serves as a bedroom, and a kitchen and bathroom. A one-bedroom apartment (1-bed) usually has one living room, one bedroom, one kitchen, and one bathroom. Sometimes, an apartment can have more than one bathroom. This usually happens when the apartment has more than one bedroom, or when the apartment is considered to be a luxury apartment.

When counting the number of rooms in an apartment, both kitchen and bathroom are usually counted as half rooms. So, a 1-bed apartment with one bathroom is a 3-room apartment. If a 1-bed has two bathrooms, or it has a dining area, it may be counted as a 3.5-room apartment. Large apartments, such as three-bedroom apartment (3-bed) or four- or more-bedroom apartment (4-bed +), may have more small rooms. A detailed description of apartment layouts in Manhattan can be found in Miller Samuel, Inc. (2002). The standard layout makes it possible for us to consider that the entire Manhattan residential real estate market is made of five sub-markets based on the number of bedrooms in each apartment. These five sub-markets are studio, 1-bed, 2-bed, 3-bed, and 4-bed +.

The Manhattan market is heavily regulated, and new apartment construction is very limited. The shortage of apartment supply increases the inefficiency of the market. However, the standard layout makes it easy to convert a larger apartment into two or more smaller apartments, or to convert two or more smaller apartments into a larger apartment. Therefore, the demand and supply in a sub-market can be adjusted even when the overall market might be in disequilibrium. This convertibility creates a lot of liquidity in the market.

The element that may introduce inefficiency lies at the market microstructure level. Manhattan is among few large cities that do not have multiple listing service. The housing information flow is not as smooth as other metropolitan areas. The supply information is exclusively controlled by brokers and it is not shared by other agents, which may block the propagation of housing information.

In almost all apartments, a maintenance fee is charged. The price of an apartment is highly leveraged by its maintenance fee. Sometimes, the apartment price is depressed by its high maintenance fee. However, the effect of a maintenance fee cannot be assessed because of the lack of relevant information.

In this study, our analysis focusses on the co-op transaction price due to the dominant quantity of co-op. The data are provided by Miller Samuel, Inc. that covers the period from the first quarter of 1989 to the fourth quarter of 2001. They are sub-market aggregated data based on 47,175 co-op transactions from more than 6,000 buildings. Those buildings are located from Downtown to the 116th Street on the west side of Central Park and to the 96th Street on the east side of Central Park. The use of sub-market aggregated data is because of the availability of these data, and also, as discussed before, the less heterogeneity of the co-ops in each sub-market.

We use the logarithm of real prices in our analysis. The real price is the deflated nominal price by the consumer price index of New York, Northern New Jersey, and Long Island, obtained from the Bureau of Labor Statistics of the U.S. Department of Labor. Table 1 presents statistics on the logarithm of real prices by sub-market.

As expected, the average prices have higher means and higher standard deviations than median prices. Average prices also have higher extreme values than median prices do in all sub-markets. The prices per room have lower means in all sub-markets. However, they have higher standard deviation compared with the other two price measures. Also, larger co-ops have higher prices per room, which indicates that the market demands a premium on the size of the co-ops. The transactions were very active in the 1-bed and 2-bed sub-markets and relatively thinner in 3-bed and 4-bed+ markets.

Figure 1 is the time series plot of the median prices. The dynamic of the median prices illustrate a bottom-out pattern. In the later 1980s, NYC experienced a real estate market bubble burst after the 1987 stock market crash. The market remained slumped until the middle of 1990s before it started to recover. Among the five sub-markets, the 4-bed+

Table 1. Statistics of co-op prices between 1989 and 2001.

Price Type	Apartment Type	Mean	Standard Deviation	Min.	Max.	Number of Transaction
Median price	Studio	6.35	0.29	5.85	6.98	6,722
	1-bed	6.98	0.24	6.67	7.48	15,989
	2-bed	7.83	0.23	7.52	8.26	19,349
	3-bed	8.80	0.24	8.47	9.34	3,463
	4-bed+	9.50	0.26	9.01	10.15	1,652
Average price	Studio	6.46	0.30	5.97	7.30	6,722
	1-bed	7.07	0.24	6.78	7.66	15,989
	2-bed	7.98	0.23	7.71	8.48	19,349
	3-bed	8.92	0.25	8.62	9.56	3,463
	4-bed+	9.75	0.30	9.21	10.70	1,652
Price per room	Studio	5.70	0.33	5.20	6.83	6,722
	1-bed	5.89	0.24	5.59	6.50	15,989
	2-bed	6.39	0.23	6.12	6.86	19,349
	3-bed	6.93	0.25	6.62	7.56	3,463
	4-bed+	7.35	0.28	6.89	8.17	1,652

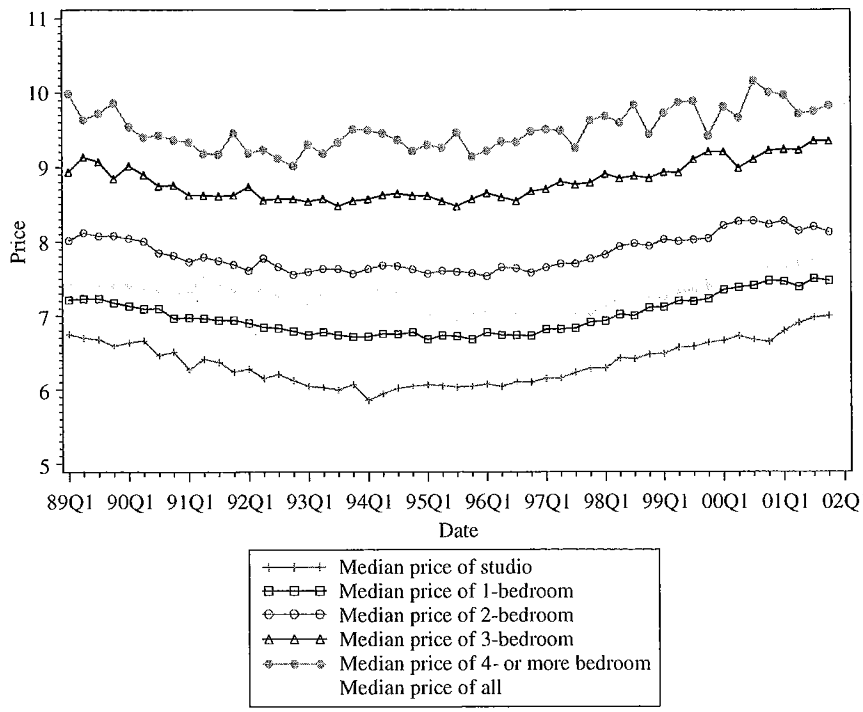

Figure 1. Time series plot for the prices of five types of co-ops.

market is shown to be the most volatile one. The time series plots for average prices and prices per room have similar patterns.

3. The model, seasonality, and stationarity

3.1. The model

Let $p_t^0, p_t^1, p_t^2, p_t^3, p_t^4$ be the logarithm of the real prices of studio, 1-bed, 2-bed, 3-bed, and 4-bed+ at time t. We use the following MVAR model to describe the co-op prices of the five sub-markets:

$$\begin{pmatrix} p_t^0 \\ p_t^1 \\ p_t^2 \\ p_t^3 \\ p_t^4 \end{pmatrix} = \delta + AD_t + \Phi_1 \begin{pmatrix} p_{t-1}^0 \\ p_{t-1}^1 \\ p_{t-1}^2 \\ p_{t-1}^3 \\ p_{t-1}^4 \end{pmatrix} + \cdots + \Phi_k \begin{pmatrix} p_{t-k}^0 \\ p_{t-k}^1 \\ p_{t-k}^2 \\ p_{t-k}^3 \\ p_{t-k}^4 \end{pmatrix} + \varepsilon_t, \tag{1}$$

where, ε_t, $t = 1, \ldots, T$, are T identically independent multivariate random variables, which are normally distributed with parameter $(\mathbf{0}, \Omega)$, $\mathbf{0}$ is a 5×1 vector of 0s and Ω is the 5×5 covariance matrix; T is the period covered by the sample, and D_t is a vector of nonstochastic variables, such as seasonal dummy variable or intervention dummy variables; δ, A, and Φ_i, $i = 1, \ldots, k$, are unknown parameters that need to be estimated. k is also unknown at the beginning and has to be determined before estimating other parameters.

The rationale behind this setup is to assume that the price of a co-op can be forecast based on its own historical price and the prices of other types of co-op in the previous quarters. If δ, A, and Φ_i are not significant in (1), then the weak efficiency of this market cannot be rejected. Otherwise, the hypothesis is rejected. If the order of autoregression $k = 1$, then the price changes follow a multivariate random walk, and the weak efficiency cannot be rejected.

When the weak efficiency hypothesis is rejected, this model allows us to estimate the interactions between prices among different sub-markets. For example, if there is a price gap between two types of property, a conversion can be done with certain cost. These relations then can be captured by the off-diagonal elements in the matrices Φ_i. Those parameters show the links among different sub-markets. By introducing the dummy variable D_t, the model also accommodates the belief that the co-op prices seasonally fluctuate. This model provides a direct method to test this assumption.

3.2. *Seasonality and stationarity*

We start our analysis by investigating the stationarity of each individual time series of co-op price. We also consider that the possible stochastic seasonality in the time series for our data has quarterly frequency. The deterministic seasonality can be modeled by introducing seasonal dummy variables. However, stochastic seasonality is related to non-stationarity of time series, and differencing has to be used at each seasonal frequency. We perform the Hylleberg et al. (1990) (HEGY) test to determine the seasonal stationarity. This test was conducted for each type of co-op price.

Specifically, assume that the prices, p_t^l, $l = 0, 1, \ldots, 4$, are generated by general autoregression processes $\phi_l(B)p_t^l - \varepsilon_t^l = 0$, where $\phi_l(B)$ is a polynomial of backward operator such that $Bp_t^l = p_{t-1}^l$ and $(1 - B^4)p_t^l = p_t^l - p_{t-4}^l$. Note that $1 - B^4$ can be decomposed into $(1 - B)(1 + B)(1 + iB)(1 - iB)$, so its unit roots are 1, -1, i, and $-i$, where $i = \sqrt{-1}$. These roots correspond to zero, semi-annual, and quarterly frequencies. Using the results in HEGY (Proposition, p. 221), the polynomial $\phi_l(B)$ can be expanded around the four roots:

$$\phi_l(B) = \pi_1^l S_1(-B) + \pi_2^l S_2(B)B + \pi_3^l S_3(B)B^2 - \pi_4^l S_3(B)(-\tilde{B}) + \phi_l^*(B)(1 - B^4),$$

where

$$S_1^l(B) = 1 + B + B^2 + B^3, \quad S_2^l(B) = 1 - B + B^2 - B^3, \quad \text{and } S_3^l(B) = 1 - B^2. \tag{2}$$

$\phi_l^*(B)$ is a polynomial of B. The test regression equation then can be written as

$$\phi_l^*(B)(1 - B^4)p_t^l = \pi_1^l S_1(B)p_t^l - \pi_2^l S_2(B)Bp_t^l - \pi_3^l S_3(B)B^2 p_t^l - \pi_4^l S_3(B)Bp_t^l. \tag{3}$$

We further consider an expanded version of (3) by adding deterministic seasonality and a time trend factor t:

$$\begin{aligned}\phi_l^*(B)(1 - B^4)p_t^l \\ = c_t^l + \pi_1^l S_1(B)p_t^l - \pi_2^l S_2(B)Bp_t^l - \pi_3^l S_3(B)B^2 p_t^l - \pi_4^l S_3(B)Bp_t^l + \varepsilon_t^l,\end{aligned} \tag{4}$$

where

$$c_t^l = \gamma_0^l + \gamma_1^l s_1 + \gamma_2^l s_2 + \gamma_3^l s_3 + \gamma_4^l t, \tag{5}$$

represents the deterministic seasonality and deterministic time trend. Equation (4), together with (5), is a regular linear regression setup. γ_i^l and π_j^l, $i = 0, \ldots, 4, j = 1, \ldots, 4$, $l = 0, \ldots, 4$, are parameters. The regressand is the co-op price, and the regressors are the transformed co-op prices by (2). Based on HEGY, the parameter π_i^l are zero when its corresponding root of $\phi_l^*(B)(1 - B^4)p_t^l$ is on the unit circle. If π_i^l is significant, we can reject the hypothesis of existing seasonal unit root. We use two t statistics to test the significance of π_1^l and π_2^l. The hypothesis of $\pi_1^l = 0$ implies a unit root at zero frequency and the hypothesis of $\pi_2^l = 0$ implies a unit root at semi-annual frequency. The alternatives for the two tests are that the parameters are smaller than the unity in an absolute sense. This implies that both tests are one-side tests. We use an F-statistic to test the joint hypothesis of $\pi_3^l = \pi_4^l = 0$, which corresponds to the existence of a unit root at quarterly frequency. The asymptotic distribution for the t-statistic and F-statistic are not the regular Student's t- and F-distributions that we usually use in regression analysis. The distributions of the t and F were derived in HEGY. Critical values are also compiled there.

Table 2 presents testing results for five different models based on choice of the terms in (5). The "No Intercept" model assumes that $\gamma_0^l, \ldots, \gamma_4^l$ are zeros. The "Constant Intercept" model assumes no seasonal dummy variables $(\gamma_1^l = \gamma_2^l = \gamma_3^l = 0)$ and no time trend $(\gamma_4^l = 0)$. The "Seasonal Dummy" model uses only seasonal dummy variables and the constant, but no time trend $(\gamma_4^l = 0)$. The "Constant Intercept and Time Trend" model does not use seasonal dummy variables $(\gamma_1^l = \gamma_2^l = \gamma_3^l = 0)$. The last model includes all terms in (5).

We perform seasonal unit roots tests for the price time series of every co-op type using five different model specifications, which give us 25 sets of testing statistics. Each set

Table 2. Testing results for seasonal unit roots.

Model	Hypothesis	Studio	1-bed	2-bed	3-bed	4-bed +
No intercept	$\pi_1^l = 0$	0.72[a]	0.52[a]	0.04[a]	0.85[a]	– 0.05[a]
	$\pi_2^l = 0$	– 3.80	– 4.38	– 3.49	– 4.53	– 2.87
	$\pi_3^l = \pi_4^l = 0$	43.00	52.36	24.59	22.90	14.01
Constant intercept	$\pi_1^l = 0$	– 0.08[a]	– 0.72[a]	– 1.06[a]	0.18[a]	– 1.47[a]
	$\pi_2^l = 0$	– 3.76	– 4.41	– 3.57	– 4.43	– 2.91
	$\pi_3^l = \pi_4^l = 0$	41.53	51.97	24.83	22.06	14.63
Seasonal dummy	$\pi_1^l = 0$	– 0.11[a]	– 0.72[a]	– 1.27[a]	0.13[a]	– 1.44[a]
	$\pi_2^l = 0$	– 3.75	– 4.29	– 2.98[a]	– 4.17	– 3.05
	$\pi_3^l = \pi_4^l = 0$	40.80	49.17	39.09	26.20	13.63
Constant intercept and linear trend	$\pi_1^l = 0$	– 0.82[a]	– 1.90[a]	– 2.42[a]	– 1.71[a]	– 3.64
	$\pi_2^l = 0$	– 3.08	– 3.52	– 3.13	– 3.84	– 2.60
	$\pi_3^l = \pi_4^l = 0$	14.46	22.96	16.44	12.81	12.58
Constant intercept, seasonal dummy and linear trend	$\pi_1^l = 0$	– 0.79[a]	– 1.86[a]	– 2.59[a]	– 1.62[a]	– 3.56[a]
	$\pi_2^l = 0$	– 3.08	– 3.38	– 2.60[a]	– 3.70	– 2.81[a]
	$\pi_3^l = \pi_4^l = 0$	14.69	21.94	28.51	15.47	11.64

[a]Cannot reject the null hypothesis of having seasonal units root at 5 percent significance level.

contains three statistics, two t-statistics, one for unit roots of zero frequency and one for semi-annual frequency, and one F-statistic for quarterly frequency. Table 2 lists all 75 ($5 \times 5 \times 3$) testing statistics for these 25 different combinations of models and co-op types. We check each statistic with its corresponding critical value in HEGY at a 5 percent significance level. We label those statistics that do not reject the null hypotheses. For example, the three statistics for the pair of model ''No Intercept'' and ''studio'' are 0.72 for hypothesis $\pi_1^0 = 0$, -3.80 for hypothesis $\pi_2^0 = 0$, and 43.00 for hypothesis $\pi_3^0 = \pi_4^0 = 0$. The second and third statistics allow us to reject the null hypotheses $\pi_2^0 = 0$ and $\pi_3^0 = \pi_4^0 = 0$ at 5 percent significance level, whereas the first statistic does not allow rejection of the null hypothesis $\pi_1^0 = 0$ at that significance level. Rejecting $\pi_2^0 = 0$ and $\pi_3^0 = \pi_4^0 = 0$ indicates that there is no seasonal unit roots at semi-annual and quarterly frequencies. But, we cannot reject the hypothesis that there exists seasonal unit roots at zero frequency because we cannot reject the hypothesis $\pi_1^0 = 0$.

The testing results in Table 2 uniformly reject the hypothesis of seasonal unit roots at semi-annual frequency and quarterly frequency, except for 2-bed and when the deterministic seasonality is introduced (the third and fifth models). The results generally indicate that the co-op prices do not have seasonal non-stationarity. However, the hypothesis at zero frequency cannot be rejected. Therefore, the co-op prices are not stationary. The standard Augmented Dickey–Fuller test also confirms that.

The results presented in Table 2 do not consider the autoregressive term of $\phi_l^*(B)(1 - B^4)p_t^l$ in (4). During our hypothesis testing, we further include as many as eight autoregressive terms in (4). The conclusions do not change from the one we draw from those based on Table 2.

Overall, time series of co-op prices are not stationary but there is no stochastic seasonality. Therefore, in the following, we need to conduct cointegration analysis to identify the relations further among the prices of different sub-markets.

4. Dynamics of property price

Although all the prices are non-stationary, it is possible to find one or more linear combinations of the prices such that the linear combinations are stationary. If these linear combinations exist, then the prices are cointegrated. The number of such linear combinations is the number of cointegration. Cointegration of prices indicates long-term equilibrium relations among those prices. To investigate relations among prices of five sub-markets, we first explore the possible cointegration among them. If there is no cointegration among these time series, the differenced time series should be used in modeling the prices. The MVAR model should be modified accordingly. If there is cointegration among them, an error correction term should be added to the model to count the short correction from long-term disequilibrium.

We use an ECM to determine the cointegration among the prices of five sub-markets. The determination of cointegration depends on the order of integration of the time series. If a time series is stationary, it is integrated of order zero (denoted as $I(0)$). If its first-order difference are stationary time series, i.e., $\Delta p_t^l = p_t^l - p_{t-1}^l$ is stationary, it is integrated of order one ($I(1)$). Similarly, we can have $I(2)$, $I(3)$, etc. We perform a series of tests to identify the order of integration. We reject the hypothesis that the co-op prices are $I(2)$. The testing results can be obtained from the author. In the following, we only report the identification of cointegration based on the prices are $I(1)$.

4.1. The error correction model

To find out the cointegrated time series, we re-write (1) in the ECM form,

$$\begin{pmatrix} \Delta p_t^0 \\ \Delta p_t^1 \\ \Delta p_t^2 \\ \Delta p_t^3 \\ \Delta p_t^4 \end{pmatrix} = \delta + AD_t + \Gamma_1 \begin{pmatrix} \Delta p_{t-1}^0 \\ \Delta p_{t-1}^1 \\ \Delta p_{t-1}^2 \\ \Delta p_{t-1}^3 \\ \Delta p_{t-1}^4 \end{pmatrix} + \cdots + \Gamma_{k-1} \begin{pmatrix} \Delta p_{t-k+1}^0 \\ \Delta p_{t-k+1}^1 \\ \Delta p_{t-k+1}^2 \\ \Delta p_{t-k+1}^3 \\ \Delta p_{t-k+1}^4 \end{pmatrix} + \Pi \begin{pmatrix} p_{t-1}^0 \\ p_{t-1}^1 \\ p_{t-1}^2 \\ p_{t-1}^3 \\ p_{t-1}^4 \end{pmatrix} + \varepsilon_t, \tag{6}$$

where, $\Delta p_{t-j+1}^l = p_{t-j+1}^l - p_{t-j}^l$, $l = 0, \ldots, 5$ and $j = 1, \ldots, k$; $\Gamma_i = \sum_{j=i+1}^k \Phi_j$, $i = 1, \ldots, k-1$; and $\Pi = I_5 - (\Phi_1 + \cdots + \Phi_k)$.

Because all the prices are $I(1)$, their differences on the left-hand side of (6) are stationary. However, the lagged price terms on the right-hand side are non-stationary. To make (6) hold, the parameter matrix Π cannot be a full rank matrix (see Johansen, 1991).

Assume that the rank of Π is r, then we can decompose matrix Π into two full rank matrices: α and β:

$$\Pi = \alpha\beta' = \begin{pmatrix} \alpha_{01} & \alpha_{02} & \cdots & \alpha_{0r} \\ \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1r} \\ \cdots & \cdots & \cdots & \cdots \\ \alpha_{41} & \alpha_{42} & \cdots & \alpha_{4r} \end{pmatrix} \begin{pmatrix} \beta_{01} & \beta_{02} & \cdots & \beta_{0r} \\ \beta_{11} & \beta_{12} & \cdots & \beta_{1r} \\ \cdots & \cdots & \cdots & \cdots \\ \beta_{41} & \beta_{42} & \cdots & \beta_{4r} \end{pmatrix}'. \tag{7}$$

Using the decomposition (7), we have

$$\Pi \begin{pmatrix} p^0_{t-1} \\ p^1_{t-1} \\ p^2_{t-1} \\ p^3_{t-1} \\ p^4_{t-1} \end{pmatrix} = \begin{bmatrix} \sum_{i=1}^{r} \alpha_{0i}(\sum_{j=0}^{4} \beta_{ji} p^j_{t-1}) \\ \sum_{i=1}^{r} \alpha_{1i}(\sum_{j=0}^{4} \beta_{ji} p^j_{t-1}) \\ \cdots \\ \sum_{i=1}^{r} \alpha_{4i}(\sum_{j=0}^{4} \beta_{ji} p^j_{t-1}) \end{bmatrix}. \tag{8}$$

In (8), the r linear combinations $\sum_{j=0}^{4} \beta_{ji} p^j_{t-1}$, $i = 1, \ldots, r$, have to be stationary. Therefore, r is the number of cointegration.

In models of (6) and (8), the r linear combinations $\sum_{j=0}^{4} \beta_{ji} p^j_{t-1}$, $i = 1, \ldots, r$, are the error correction terms. The terms involving differences of prices and dummy variables are autoregressive terms. The last term is the random error term. It is possible that the error correction term also contains a constant term and a time trend term. There is a difference between models that have a constant or time trend term in the error correction term and those that have them in the autoregressive term. The constant and time trend in the error correction term indicates the long-term relation between prices and time.

The model (6) and (7) has very rich contents to explain the dynamics of the co-op prices. The autoregressive term contains a possible constant (δ), a set of dummy variables (D_t), and previous price changes (Δp^l_{t-i}). The error correction term is a linear combination of previous prices ($\sum_{j=0}^{4} \beta_{ji} p^j_{t-1}$) adjusted by the adjustment speed parameters (α_{ij}). It is direct to see that A and $\Gamma_1, \ldots, \Gamma_{k-1}$ allow us to estimate the impact from dummy variables, such as deterministic seasonality, and that from previous price changes. However, the role of α_{ij}, β_{ij}, and the error correction terms are a little subtle.

To be clear, let us take the studio price changes (Δp^0_t) as our example. Based on the explanation in Engle and Granger (1987), equation $\sum_{j=0}^{4} \beta_{ji} p^j_{t-1} = 0$ describes the long-term relationship among the co-op prices in equilibrium. Therefore, $\sum_{j=0}^{4} \beta_{ji} p^j_{t-1}$ is stationary ($I(0)$) if there exists a cointegration among prices p^j_{t-1}, $j = 0, \ldots, 4$. If $\sum_{j=0}^{4} \beta_{ji} p^j_{t-1} \neq 0$, then the prices in the last period are deviated from the equilibrium prices. For a market with a self-adjustment mechanism, the price change from the last period should include a term to reflect the correction of this deviation, which is $\sum_{i=1}^{r} \alpha_{0i}(\sum_{j=0}^{4} \beta_{ji} p^j_{t-1})$ in our model. This correction term summarizes price deviations from more than one long-term relationship between co-op prices in different sub-markets. We further model the proportion of deviation used in the correction term, which is described by the parameter α_{0j}. If α_{0j} is zero, then there is no correction come from the jth long-term relationship if the price are deviated from their equilibrium price. If α_{0j} is large than one, the correction may be over done. Overall, α_{ij} provide adjustment speeds of the prices from disequilibrium to equilibrium and β_{ij} reveal the long-term relationships among prices from all sub-markets when the markets are in equilibrium.

Model (6) can be further extended to include a time trend term on the right-hand side of the equation. Adding this term implies that the price changes have a component that is linearly related to time. A constant term, dummy variables, or a time trend may also be added to the long-term relationship. With those terms, the model takes into account a possible seasonality and time trend in the short-term.

4.2. *Model identification and estimation*

We use the maximum-likelihood estimation proposed by Johansen (1991) to estimate the parameters in (6) and (8). The log-likelihood function, given the sample p_t^l, $l = 0, \ldots, 4$ and $t = 1, \ldots, T$, is

$$\begin{aligned} &L(\delta, A, \Gamma_1, \ldots, \Gamma_{k-1}, \alpha, \beta, \Omega; p_t^l, t = 1, \ldots, T, l = 0, \ldots, 4) \\ &= -\frac{T-1}{2}\log(2\pi) + \frac{1}{2}\log(|\Omega^{-1}|) - \frac{1}{2(T-1)}\sum_{t=1}^{T-1} R_t'\Omega^{-1}R_t, \end{aligned} \tag{9}$$

where

$$\begin{aligned} R_t = &\begin{pmatrix} \Delta p_t^0 \\ \Delta p_t^1 \\ \Delta p_t^2 \\ \Delta p_t^3 \\ \Delta p_t^4 \end{pmatrix} - \delta - AD_t - \Gamma_1 \begin{pmatrix} \Delta p_{t-1}^0 \\ \Delta p_{t-1}^1 \\ \Delta p_{t-1}^2 \\ \Delta p_{t-1}^3 \\ \Delta p_{t-1}^4 \end{pmatrix} - \cdots - \Gamma_{k-1} \begin{pmatrix} \Delta p_{t-k+1}^0 \\ \Delta p_{t-k+1}^1 \\ \Delta p_{t-k+1}^2 \\ \Delta p_{t-k+1}^3 \\ \Delta p_{t-k+1}^4 \end{pmatrix} \\ &- \begin{pmatrix} \alpha_{01} & \alpha_{02} & \cdots & \alpha_{0r} \\ \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1r} \\ \cdots & \cdots & \cdots & \cdots \\ \alpha_{41} & \alpha_{42} & \cdots & \alpha_{4r} \end{pmatrix} \begin{pmatrix} \beta_{01} & \beta_{02} & \cdots & \beta_{0r} \\ \beta_{11} & \beta_{12} & \cdots & \beta_{1r} \\ \cdots & \cdots & \cdots & \cdots \\ \beta_{41} & \beta_{42} & \cdots & \beta_{4r} \end{pmatrix}' \begin{pmatrix} p_{t-1}^0 \\ p_{t-1}^1 \\ p_{t-1}^2 \\ p_{t-1}^3 \\ p_{t-1}^4 \end{pmatrix}. \end{aligned} \tag{10}$$

The estimates of $\delta, A, \Gamma_1, \ldots, \Gamma_{k-1}, \alpha, \beta$, and Ω are those that maximize the log-likelihood function $L(\delta, A, \Gamma_1, \ldots, \Gamma_{k-1}, \alpha, \beta, \Omega\, p_t^l, t = 1, \ldots, T, l = 0, \ldots, 4)$:

$$L(\hat{\delta},\hat{A},\hat{\Gamma}_1,\ldots,\hat{\Gamma}_{k-1},\hat{\alpha},\hat{\beta},\hat{\Omega};p_t^l,t=1,\ldots,T,l=0,\ldots,4)$$
$$= \max_{\substack{\delta,A,\\ \Gamma_1,\ldots,\Gamma_{k-1},\\ \alpha,\beta,\Omega}} L(\delta,\Psi,\Gamma_1,\ldots,\Gamma_{k-1},\alpha,\beta,\Omega;p_t^l,t = 1,\ldots,T,l = 0,\ldots,4).$$

We start with the model estimation by identifying the order of the autoregressive term in (6). We use a set of goodness-of-fit measures and information criteria in determining the order. We calculated the determinant of the estimated covariance matrix $|\hat{\Omega}|$, which is equivalent to the estimated variance of the standard error in the univariate time series. As we expect, when more parameters are added to the model, the fit of the model gets better and the $|\hat{\Omega}|$ gets smaller. Although this measure does not provide information for determining the order of autoregression, it does provide information that the algorithm performs well. The information criteria used here are the famous Schwarz Bayesian Criterion (SBC) (see Hannan and Deistler, 1988) and Hanna–Quinn Criterion (HQC) (see Quinn, 1980). Both SBC and HQC penalize adding more parameters to the model. We also perform the Ljung–Box (LB) test and two Lagrange Multiple (LM) type tests to test the significance of autocorrelation among residuals from models with different autoregressive terms. The LB and LM tests are described in Ljung and box (1978) and Godfrey (1988), respectively. We prefer those models that eliminate the autocorrelation in the residuals.

Because different model setups lead to different asymptotic distributions of the testing statistics, our investigation is very comprehensive and include all the model setups discussed in the previous section. We eliminate models that contain time trends. Figure 2 presents the time series plots for the five types of co-op price changes. We do not observe any time trend in the price changes. Formal statistical tests using time trends in the model also confirm that. Figure 1 also indicates that the price time series do not have long-term time trend. We perform tedious statistical tests to confirm that. The results for these tests can be obtained from the author. Therefore, in the following, we only report the results based on three model setups: there is no constant in either the autoregressive term or the error correction term, there is only a constant in the error correction term and there is a constant only in the autoregressive term.

Table 3 reports testing results for these three models. Because the data have quarterly frequency, with each model, we suspect that the order of autoregression k can be from 1 to 4. For each selected order of autoregression, there are up to four possible cointegrations among these five price time series. There is a total of 48 ($4 \times 4 \times 3$) sets of testing statistics. We estimate the parameters in the model and calculated five statistics in each case. It is direct to see that, for all three models, the $|\hat{\Omega}|$ demonstrates monotonic decreasing as expected. The SBC and HQC obtain their lowest levels when the order of autoregression is 2 and increase after that. The autocorrelation tests LM(2) and LM(4) reject the hypothesis of no autocorrelation for order one models, but they do not reject the hypotheses for models of order two, three, or four. The results indicate that 2 is a reasonable choice for the order of autoregressive based on the data we use. We perform similar analysis described above by adding back deterministic seasonal dummy variables. We also examine the significance of the included dummy variables. The order of autoregression is the same as before. The standard t-test indicates that the deterministic

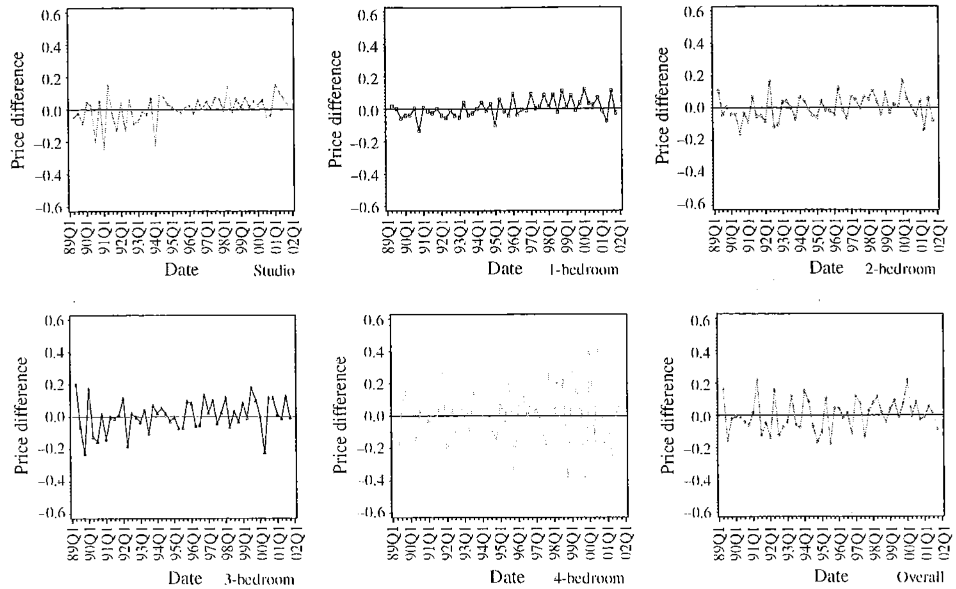

Figure 2. Time series for the returns of five types of co-op.

seasonal dummy variables are not significant. So, we do not include the deterministic seasonal dummy variable in our analysis. In the following, the analysis is based on (6) with $k = 2$.

We determine the number of cointegration based on the trace test proposed by Johansen (1988). In maximizing the log-likelihood function (9), as derived by Johansen (1988), the r vectors of β in (8) can be expressed as r eigenvectors with r eigenvalues: λ_i, $i = 1, \ldots, r$. Intuitively, the eigenvalues λ_is are the squared correlation coefficients between the linear combination of the prices and the linear combination of the differences of prices. If the linear combination of prices is non-stationary, the correlation between these two linear combinations should be small and the eigenvalues should be close to zero. Formally, assuming that the null hypothesis H_0 is $\lambda_{r+1} = \cdots = \lambda_5 = 0$, then the likelihood ratio test statistic (trace statistic) for the hypothesis is $\hat{\Lambda}_r = -T\sum_{i=r+1}^{5}\log(1 - \hat{\lambda}_i)$, $r = 0, \ldots, 4$. Johansen and Nielsen (1993) derived the asymptotic distributions for the test statistics. Table 4 presents trace statistics for three possible models.

To determine the number of cointegrations using the statistics in Table 4, we use a sequence of hypothesis testings, starting with the hypothesis of 5 unit roots ($r = 0$). If this hypothesis is rejected, it implies that $\lambda_1 > 0$. Then, we continue to test the hypothesis $\lambda_2 = \cdots = \lambda_5 = 0$. When the hypothesis cannot be rejected, we obtain the number of the unit roots. The results in Table 4 indicates that when Model 2, 3 and 4 are true, the null hypotheses of $r = 1$ is rejected but $r = 2$ cannot be rejected. This implies that, under these models, $r = 2$ is a proper choice. When Model 1 is the correct one, we have $r = 1$.

Table 3. Results for choosing autoregressive terms.

Models	Statistics	Autoregression Order $k=1$				Autoregression Order $k=2$				Autoregression Order $k=3$				Autoregression Order $k=4$			
		$r=1$	$r=2$	$r=3$	$r=4$	$r=1$	$r=2$	$r=3$	$r=4$	$r=1$	$r=2$	$r=3$	$r=4$	$r=1$	$r=2$	$r=3$	$r=4$
1	$\lvert\hat{\Omega}\rvert$	−26.47	−25.40	−25.95	−26.37	−27.61	−26.97	−27.40	−27.58	−28.76	−28.10	−28.53	−28.74	−30.06	−29.35	−29.75	−30.02
	SBC	−24.54	−24.71	−24.71	−24.75	−23.70	−24.31	−24.19	−23.98	−22.80	−23.42	−23.28	−23.10	−21.99	−22.57	−22.41	−22.28
	HQC	−25.12	−24.92	−25.09	−25.25	−24.88	−25.12	−25.16	−25.07	−24.60	−24.83	−24.87	−24.80	−24.42	−24.61	−24.62	−24.61
	LM (2)(%)	0	0	0	0	7	1	6	6	5	21	6	3	59	29	22	54
	LM (4)(%)	0	0	1	1	14	8	7	11	29	62	26	20	27	49	40	28
2	$\lvert\hat{\Omega}\rvert$	−27.02	−25.48	−26.14	−26.68	−28.13	−26.99	−27.62	−27.95	−29.07	−28.12	−28.58	−28.89	−30.52	−29.35	−29.88	−30.27
	SBC	−24.71	−24.70	−24.76	−24.83	−23.83	−24.25	−24.25	−24.12	−22.71	−23.35	−23.18	−23.02	−22.05	−22.50	−22.38	−22.29
	HQC	−25.41	−24.94	−25.18	−25.39	−25.13	−25.08	−25.27	−25.28	−24.63	−24.79	−24.81	−24.79	−24.60	−24.56	−24.64	−22.29
	LM (2)(%)	1	0	0	0	28	6	12	32	2	26	11	2	13	30	16	9
	LM (4)(%)	5	1	4	11	18	7	17	19	18	69	36	20	9	48	41	14
3	$\lvert\hat{\Omega}\rvert$	−27.02	−25.49	−26.15	−26.69	−28.13	−27.01	−27.63	−27.96	−29.07	−28.14	−28.60	−28.91	−30.52	−29.39	−29.92	−30.31
	SBC	−24.71	−24.41	−24.54	−24.68	−23.83	−23.95	−24.03	−23.96	−22.71	−23.06	−22.96	−22.87	−22.05	22.21	−22.18	−22.16
	HQC	−25.41	−24.74	−25.03	−25.29	−25.13	−24.88	−25.12	−25.17	−24.63	−24.59	−24.66	−24.69	−24.60	−24.37	−24.51	−24.61
	LM (2)(%)	1	0	0	0	28	6	13	32	2	24	12	2	15	24	11	7
	LM (4)(%)	5	1	5	14	19	9	21	22	20	73	41	23	10	51	44	13
4	$\lvert\hat{\Omega}\rvert$	−27.50	−25.52	−26.20	−26.82	−28.94	−27.08	−28.01	−28.60	−29.77	−28.20%	−28.91	−29.36	−30.52	−29.39	−29.92	−30.31
	SBC	−24.80	−24.36	−24.43	−24.59	−24.25	−23.95	−24.26	−24.38	−23.02	−23.04	−23.11	−23.09	−22.05	−22.21	−22.18	−22.16
	HQC	−25.62	−24.71	−24.97	−25.27	−25.67	−24.90	−25.39	−25.66	−23.02	−24.59	−24.86	−24.98	−24.60	−24.37	−24.51	−24.61
	LM (2)(%)	0	0	0	0	33	29	40	44	16	60	25	44	15	24	11	7
	LM (4)(%)	9	1	3	8	15	5	19	31	5	60	49	29	10	51	44	13

Table 4. Trace statistics in determining the number of cointegration.

		Models		
H_0: Rank $= r$	H_a: Rank $> r$	1	2	3
0	0	77.78[a]	102.90[a]	103.62[a]
1	1	32.06	56.12[a]	56.79[a]
2	2	10.64	24.89	25.56
3	3	1.63	8.66	8.84
4	4	0.09	0.88	0.99

[a]Reject the null hypothesis under level of significance 0.05.

Note that Model 1 is a constrained version of Models 2 and 3 when the constants in both models are set to zeros. Also, Model 2 is a constrained version of Model 3 when the constant is only allowed in the error correction term. Therefore, we can perform statistical tests to validate the significance of those constraints. We use the χ^2 statistics in Johansen (1991) to carry out the tests. The tests rejected those constraints imposed on Model 1 at a significance level of 0.01 and cannot reject the one imposed by Model 2. Therefore, Model 2 was chosen. Given Model 2, the number of cointegration $r = 2$. After normalizing the estimated β, the α, β, and Π in Model 2 can be estimated as:

$$
\begin{aligned}
\hat{\Pi} &= \hat{\alpha}\hat{\beta}' \\
&= \begin{pmatrix} 0.315 & -0.295 \\ -0.152 & -0.111 \\ 0.203 & -0.232 \\ 0.778 & 0.103 \\ -0.951 & 0.079 \end{pmatrix} \begin{pmatrix} 0.017 & 1.000 & -1.203 & -0.424 & 0.553 & 0.809 \\ 1.000 & -0.047 & 0.578 & -1.862 & 0.020 & 5.640 \end{pmatrix} \\
&= \begin{pmatrix} -0.290 & 0.329 & -0.550 & 0.417 & 0.168 & -1.412 \\ -0.114 & -0.146 & 0.118 & 0.272 & -0.086 & -0.751 \\ -0.229 & 0.214 & -0.379 & 0.347 & 0.108 & -1.147 \\ 0.116 & 0.773 & -0.876 & -0.521 & 0.432 & 1.208 \\ 0.063 & -0.954 & 1.190 & 0.255 & -0.524 & -0.322 \end{pmatrix}.
\end{aligned} \tag{11}
$$

Note that in (11), the estimated β are not unique. For easy comparison, we normalize the estimated β by the 1-bed price for the first cointegration and the studio price for the second cointegration.

Further study on the estimate of (11) revealed the following relationship, $p_t^2 - p_t^1 = 0.5(p_t^4 - p_t^3) + d_1$ and $p_t^0 + p_t^2 = 2p_t^3 + d_2$, where d_1 and d_2 are two constants. The first states that the price gap change between 2-bed and 1-bed is about half that of 4-bed + and 3-bed. The second indicates that the price change of 3-bed is the average of the price changes of 1-bed and 2-bed. The details of the study can be provided upon request.

4.3. Weak exogeneity

We have established an order 2 MVAR model (1), and, equivalently, its ECM form (6). It seems that we can conclude that the weak efficiency assumption should be rejected by our findings. However, in the MVAR model, it is possible that one or more variables weakly exogenous exist. The existence of weak exogeneity may alter the model specification and therefore void the weakly inefficiency conclusion.

We suspect that at least one weak exogenous variable exists, which is evidenced by smaller values of $\alpha_{21} = -0.152$ and $\alpha_{22} = -0.111$ in (11). We conjecture that these two parameters may not be significant. If they are not significant, then the disequilibrium of the long-term relation has no influence on 1-bed. However, the price of 1-bed influences the prices of other co-ops. In other words, the 1-bed may be weakly exogenous.

We use a likelihood ratio test proposed in Johansen and Juselius (1992) to test the weak exogeneity for all five prices. The statistic has a χ^2 distribution. Under significance level 0.05, the test rejects the weak exogeneity hypothesis of studio, 2-bed, and 3-bed, but cannot reject the weak exogeneity of 1-bed and 4-bed +, where 1-bed has higer p-value. We refit the model using four time series: studio, 2-bed, 3-bed, and 4-bed +. Then, we test the weak exogeneity of each of the prices remained in the model (studio, 2-bed, 3-bed, and 4-bed +). All tests reject the hypotheses of weak exogeneity. 1-bed is the only weakly exogenous variable. After considering the weak exogeneity of 1-bed, we re-estimate the parameters in the model. We check the significance of all parameters in the model and the model setup. Specifically, we check the estimated α and their corresponding t-statistics. The results are presented in Table 5. Note that α_{21} and α_{42} are not significant. However, 2-bed and 4-bed + are still significant because α_{22} and α_{41} are significant.

The value of α provides information on the speed of adjustment when there is a long-term disequilibrium. For the 4-bed + margin premium disequilibrium, 3-bed and 4-bed + adjust much slower than do studio and 2-bed. The former takes about a quarter to adjust, whereas the later takes only a month. For the 3-bed undervalue disequilibrium, the adjustments are almost the same for all relevant apartment, taking less than a month. Overall, this market demonstrates very fast adjustment to the prices to achieve equilibrium among different types of apartments.

Table 5. Estimated adjustments to disequilibrium.

Variables	Estimate	t-value
α_{01}	− 0.429	− 2.332
α_{02}	− 0.211	− 2.932
α_{21}	− 0.306	− 2.038
α_{22}	− 0.162	− 2.760
α_{31}	− 1.081	− 5.797
α_{32}	0.222	3.045
α_{41}	1.057	2.449
α_{42}	0.096	0.571

In the end, we estimate the final model, with only significant coefficients:

$$\begin{aligned}
\Delta p_t^0 &= -0.211p_t^0 - 0.591p_t^2 + 0.314p_t^3 + 0.206p_t^4 + 0.333p_t^1 - 1.067,\\
\Delta p_t^2 &= -0.162p_t^0 - 0.431p_t^2 + 0.147p_t^4 + 0.237p_t^1 - 0.829,\\
\Delta p_t^3 &= 0.222p_t^0 - 0.910p_t^2 - 0.771p_t^3 + 0.519p_t^4 + 0.838p_t^1 + 1.718,\\
\Delta p_t^4 &= 1.131p_t^2 - 0.508p_t^4 - 0.820p_t^1.
\end{aligned} \tag{12}$$

To be sure, we examine the autocorrelation in the residuals. We perform the LB test and the two LM tests used in Table 3. These tests cannot reject the hypotheses that the autocorrelations are zero with p-values 0.08, 0.16, and 0.36. We further use a method by Doornik and Hansen (1994) to test the normality of the residuals obtained from the model. The hypothesis of normality cannot be rejected with p-value 0.24. Finally, we check the module of the eigenvalues of our final model. The highest value is 0.6310, which is well below 1. When the absolute values of all the eigenvalues are less than 1, the estimated model (12) will converge to a equilibrium state. By now, our model is identified and estimated.

As for the goal of this paper, (12) indicate that prices and the returns of co-ops can be predicted from their previous values. This indicates that the weak efficiency hypothesis should be rejected.

5. Conclusion and implications

Using Manhattan residential co-op median prices from 1989 to 2001, we developed an MVAR model to investigate the weak efficiency of this market. To avoid possible model misspecification, which may lead to false conclusions, we carefully examined a variety of model setups. We consider stationarity and seasonal stationarity of the prices, deterministic seasonality, cointegration with and without time trend, seasonality and constant, and several possible autoregressive terms in our MVAR. We also investigated the impact of co-op price's weak exogeneity on the model specification. We identified an order 2 MVAR model to describe the dynamics of the co-ops prices. Our finding essentially rejected the weak efficiency of the Manhattan co-op market. This result also held when mean price and price per room were used in our research, which enhanced the conclusion that the Manhattan co-op market is not weakly efficient.

Recall that the co-ops within each sub-market are relatively homogeneous; the transactions are very active and the transaction information is well observed by market participants; and the commission fee in Manhattan is relatively lower. We therefore infer that the cause of the inefficiency of the market may not be the heterogeneity of the properties, the propagation of transaction information or explicit expenses incurred in the transaction. Capozza et al. (2002) explored two hypotheses for prices serial correlation: information explanation and supply-based explanation. They used population and real income as proxies of information cost. The data in our study confirm their approach of using population and real income as proxies of information cost in considering that Manhattan is a very populous city with higher income home buyers. Therefore, we believe that the inefficiency may come from the supply side. We also suspect that market microstructure and implicit transaction cost also play important roles in weak efficiency.

We have observed elements that may cause inefficiency in the Manhattan market. First, on the supply side, the Manhattan market has been the tightest market among all metropolitan markets. Meanwhile, the Manhattan market is highly regulated. The notorious zoning policy and approval procedure make it very costly for developers to supply additional apartments in response to demand increases (see Salins, 2002; Salama et al. 2002). Second, the market microstructure in Manhattan is very special. Besides the lack of a multiple listing service, the brokers' non-cooperative attitude also blocks the demand and supply information flow. There are a large number of brokers working in Manhattan. However, when a broker is approached by a seller to sell a co-op, the broker is motivated to share this information with other brokers, who may know some potential buyers. Instead, the broker always tries to find a buyer directly and get all the commission. This search process may take very long in comparison with a process that reveals the co-op in public domain and promptly finds the buyers. In contrast, brokers know that a buyer may also want to find the seller directly. Third, the apartments in our study are co-ops. The majority apartments in Manhattan are co-ops, where buyers actually buy a certain amount of shares of a cooperation that owns the building and get the right to live in the building. Usually, the board of the cooperation has a strict review process in place to examine the qualification of the potential buyer. The board can reject the buyer without any reason. This approval procedure decreases the possibility of making a deal. Although these elements may help us to explain the inefficiency, further empirical research is needed to confirm the role they play.

Besides prompting further questing the source of inefficiency, our research discovers two long-term relationships among prices of different sub-markets. We also identified a sub-market, 1-bed, that tends to feed information to the rest of other sub-markets. This sub-market has the highest transaction volume and is the type of co-op that most buyers first purchase when they decide to own an apartment in Manhattan. This sub-market shows weak exogeneity when median price and price per room are used in modeling. More research is needed to understand whether the weakly exogeneous variable is related to the high transaction volume.

Acknowledgments

This research was done when the auther was at Baruch College and it was supported by Eugene Lang Junior Faculty Research Fellowship. All opinions expressed in this article are the author's own and do not necessarily reflect the opinions of Deutsche Bank and its affiliates. The author is grateful to two anonymous referees for their detailed comments, suggestions, critics and encourages. The author thanks Janathan J. Miller, President of Miller Samuel Inc., who kindly provided the data for this research. The author also thanks

Steve Hanson and Don McBride from Cadmus, and his neighbor Ruth Ohman for their help with the presentation.

References

Abraham, J. M., and P. H. Hendershott. (1996). "Bubbles in Metropolitan Housing Market," *Journal of Housing Research* 7(2), 191–208.

Capozza, D. R., P. H. Hendershott, C. Mach, and C. J. Mayer. (2002). "Determinants of Real House Price Dynamics." Technical Report, National Bureau of Economic Research, Cambridge, Massachusetts.

Cho, M. (1996). "Housing Price Dynamics: A Survey of Theoretical and Empirical Issues," *Journal of Housing Research* 7(2), 145–172.

Doornik, J. A., and H. Hansen. (1994). "An Omnibus Test for Univariate and Multivariate Normality." Technical Report, Nuffield College, Oxford.

Engle, R. F., and C. W. J. Granger. (1987). "Co-intergration and Error Correction: Representation, Estimation, and Testing," *Econometrica* 55, 251–276.

Fama, E. F. (1970). "Efficient Capital Markets: A Review of Theory and Empirical Work," *Journal of Finance* 25, 383–420.

Godfrey, L. G. (1988). *Misspecification Tests in Econometrics: The Lagrange Multiplier Principle and Other Approaches*. Cambridge: Cambridge University Press.

Hannan, E. J., and M. Deistler. (1988). *The Statistical Theory of Linear Systems*. New York: John Wiley & Sons.

Hylleberg, S., R. Engle, C. Granger, and B. Yoo. (1990). "Seasonal Intergration and Cointegration," *Journal of Econometrics* 44, 215–238.

Johansen, S. (1988). "Statistical Analysis of Cointegration Vectors," *Journal of Economic Dynamics and Control* 12, 231–254.

Johansen, S. (1991). "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models," *Journal of Econometrics* 52, 389–402.

Johansen, S., and K. Juselius. (1992). "Testing Structural Hypotheses in a Multivariate Cointegration Analysis of the PPP and the UIP for U.K.," *Oxford Bulletin of Economics and Statistics* 52, 169–210.

Johansen, S., and B. Nielsen. (1993). *Asymptotics for Cointegration Rank Tests in the Presence of Intervention Dummies*. Institute of Mathematical Statistics, University of Copenhagen.

Ljung, S., and G. Box. (1978). "On a Measure of Lack of Fit in Time Series Models," *Biometrika* 65, 297–303.

Malpezzi, S. (1999). "A Simple Error Correction Model of Housing Prices," *Journal of Housing Economics* 8, 27–62.

Meese, R., and N. Wallace. (1993). "Residential Housing Prices in the San Francisco Bay Area: New Test of Explanatory Power of Economic Fundamentals," Technical Report, Hass School of Business, University of Calfornia, Berkeley.

Miller Samuel, Inc. (2002). *Manhattan Residential Real Estate Market, 2002*. New York City, New York.

Quinn, B. G. (1980). "Order Determination for a Multivariate Autoregression," *Journal of the Royal Statistical Society, B* 42, 182–185.

Salama, J. J., M. H. Schill, and M. E. Stark. (2002). *Reducing the Cost of New Housing Construction in New York City*. School of Law, New York University.

Salins, P. D. (2002). "New York City's Housing Gap, Civic Report 2." New York: Manhattan Institute for Policy Research.

Zhou, Z. (1997). "Forecasting Sales and Price for Existing Single-Family Homes: A VAR Model with Error Correction," *Journal of Real Estate Research* 14(1/2), 155–167.