Structural Change, Cointegration, and Market Clearing

In the previous section, the retail and farm price equations of 1 derive from market-clearing conditions for consumer food products and for farm ingredients (see Appendix). However, questions concerning the specification of equations 1 arise if markets have undergone structural change.

Regardless of the impact of structural change, market clearing means that excess supply (demand) would be of short duration and would equal zero on average. Furthermore, the effects of unforeseen shocks to market clearing would die out over time. In time series jargon, excess supply (demand) would be stationary. It is straightforward to show (using the market-clearing conditions shown in the Appendix) that the error terms, ε_{rj} and ε_{fj} , of equations 1 may represent excess supply (demand) variables for food and farm products. Stationary error terms imply market clearing, and market clearing would prevent the variables of equations 1 from moving too far apart.

Associated with structural change are market trends. Evidence of trends or changes in trends of market variables are often used as indicators of structural change. In time series jargon, variables that are characterized or generated by trends are non-stationary. However, even if each of the variables of equations 1 displays trends, the equations would still reflect market clearing if the excess supply variables or error terms are stationary. If each of the time series variables used in a model displays trends, and if model errors are stationary, the model is *cointegrated* (Engle and Granger). Tests of cointegration are tests of whether the data support the theory. In a cointegrated regression, some mechanism cancels or aligns the trends among the variables, and in equations 1 the mechanism is market clearing.

If the trends driving the non-stationary variables of equations 1 are not linked, excess supply would not die out, and the regression would be *spurious* (Granger and Newbold, p. 202-214, Hamilton p. 557-562). If the price equations of (1) are spurious, inter-temporal movements in one set of variables of equations 1 do not explain inter-temporal movements in other variables. A finding of a spurious regression would not support the theory.

In general, if market variables are driven by trends, the trends can be *deterministic*, stochastic, or a combination of both. A deterministic trend is a time trend defined in the usual way. Trends in demographics or predictable increases in real wages and productivity over the last century may drive the deterministic portion of trends in market variables. Market variables generated by deterministic trends pose few problems for statistical inference because with an infinite number of observations, such variables can be forecast from past observations with an arbitrary degree of accuracy. The second type of trend is a *stochastic* trend. Variables driven by stochastic trends are referred to as *unit root* or *integrated* series. For example, trends in real wages tied to unpredictable changes in the direction of inflation, unpredictable changes in the direction of consumer demand, technology, or the continual process of industrial reorganization, may be generating stochastic trends in market variables.[11] Unlike deterministic trends, stochastic trends change direction unpredictably. Integrated market variables pose special problems for statistical inference because even in infinite samples, optimal forecasts of these variables do not converge, but are continually revised as new observations become available.

More formally, the accuracy and reliability of forecasts of market variables depend on whether the variable is driven by a deterministic or a stochastic trend. As the forecast horizon grows, the forecast of a series generated by a deterministic trend converges to a time trend, and the mean squared error (MSE) of this forecast converges to the unconditional variance of the series (Hamilton, p. 438-42). Population, real wages, and real disposable income may be accurately and reliably forecast. On the other hand, tastes and preferences, technology, and the continual reorganization of an industry may be stochastic trends because changes in any of these may be impossible to predict. Unlike deterministically trending variables, the forecast of a unit root variable diverges with the length of the forecast horizon, and the MSE of the forecast increases without bound (Hamilton, p. 438-42).

Associated with each type of trend is a type of cointegration. A model constructed from deterministically trending variable series is *deterministically cointegrated* if the deterministic trends in the model's variables are linked. In practice, a regression model is

[11] That is, a series that is stationary around a deterministic trend.

deterministically cointegrated if a time trend variable appended to the model is not statistically different from zero. A model constructed from a set of stochastically trending series is *stochastically cointegrated* if the model errors are stationary. Just as market variables may reflect both a deterministic and a stochastic trend, a model may be both deterministically and stochastically cointegrated.

Using annual time series from 1958-97, we computed Dickey-Fuller and Phillips-Perron t-tests for the logged and deflated variables used in the seven sets of retail and farm price equations.[12] Both sets of tests are designed to refute the hypothesis that, conditioned on an AR(1) representation, a single unit root net of an intercept (or drift) or net of a deterministic time trend governs the series. The tests differ in the way they handle serial correlation of the error terms of the AR(1) specification.[13] Almost without exception, the two sets of tests suggest that both a stochastic and a deterministic trend drive most of the variable series used in equations 1.[14]

Given evidence of trends in the variables, the question is whether these equations are stochastically and deterministically cointegrated. The specification of equations 1 used throughout this report is as follows. The deterministic regressors include an intercept and a deterministic time trend. The stochastic regressors include a vector of (logged) nonfarm input prices ($\ln \mathbf{W}$), a (log) farm supply variable specific to market i ($\ln Q_i$), and the total demand shifters ($\ln Z_i$) for each of the consumer demand equations.[15] The vector $\ln \mathbf{W}$ consists of (logs of) wages, the price of packaging, the price of transportation, and the price of energy. To satisfy homogeneity, all prices and income variables in equations 1 and 2 are deflated by the price of other nonfarm inputs (Elitzak). Hence, the tests for cointegration are based on a specification that includes six stochastic regressors, a constant, and a deterministic time trend for each retail and farm price equation. We compute two sets of tests.[16]

The first is based on model residuals, and specifically tests the null hypothesis of a spurious regression. Again, a model is spurious (or not *stochastically* cointegrated) if the model errors follow a unit process. Engle and Granger (1987) suggest applying Dickey-Fuller tests to model residuals. Phillips and Ouliaris confirm that the Dickey-Fuller and Phillips-Perron statistics can be used to test for spurious regressions. They find, however, that the critical values depend not on the number of observations, but on the number of stochastic regressors used in model specification, and whether the regression includes an intercept or a deterministic time trend.

Table 1 reports Dickey-Fuller and Phillips-Perron t-tests designed to refute the null that the equations 1 are spurious regressions. The statistics are based on Ordinary Least Squares (OLS) residuals. The Dickey-Fuller results presented in table 1 fail to reject the null of a spurious regression for each of the 14 equations, while the Phillips-Perron results fail to reject the null for 8 of the 14 equations.

[12] The sample series used to create the variables are discussed in the Appendix. All price and income variables are deflated by the price of other nonfarm inputs and logged prior to testing. The total demand shifters are computed directly from the estimates of the double-log level system of consumer demand equations. Prices and income are also deflated by the price of other nonfarm inputs prior to estimation. The test results were computed using Shazam.

[13] Dickey and Fuller control for the serial correlation in the error terms by adding lags to the AR(1) representation. Phillips and Perron compute estimates of the covariogram of the errors of an AR(1) process. For a concise comparison between the tests we refer the reader to Hamilton (p. 504-518). The simulations presented in Phillips and Perron (1988) reveal that neither test is universally more powerful than the other.

[14] The Dickey-Fuller and Phillips-Perron results are available upon request. Some results conflict. For example, the results suggest the farm prices for poultry and eggs may be stationary around a time trend (-3.95 and -4.49) but are unit root non-stationary around an intercept. The results also suggest the farm supply for beef may be stationary around a constant but non-stationary around a time trend. When unit root tests conflict, Holden and Perman spell out a multi-step procedure that may be useful in sorting out the results.

[15] The seven-equation demand system was also found to be stochastically cointegrated, and seemingly unrelated canonical cointegrating regression estimates with an intercept and no time trend and symmetry and homogeneity imposed are used to construct the demand shift variables used in equations 1.

[16] Because we found that a linear deterministic time trend variable was invariably statistically different from zero, we reject the null of deterministic cointegration for each price equation and included it in the model specifications for each industry. Hence, our tests of cointegration are tests specifically for stochastic cointegration.

The second set of tests is designed to examine the null that the regressions are stochastically cointegrated. The tests are based on the observation by Park (1990) that appending a set of integrated series to a stochastically cointegrated model yields a spurious regression. If the additional variables add no explanatory power to the regression, they are superfluous, and the original model specification is cointegrated. The technical problem of testing whether the variables are superfluous is that, in general, the model error terms are correlated with the first differences of the model's regressors. This correlation destroys the asymptotic normality of parameter estimates, and hence destroys the reliability of the usual chi-square tests. Park (1990) derives a transformation that accounts for this correlation, and uses it to transform each of the variables of a model. Chi-square tests based on the transformed regression represent valid tests of the null that the additional variables are superfluous, and the original model is stochastically cointegrated.

Table 1 — Residual-based tests of spurious regressions

	Dickey-Fuller	Phillips-Perron
Retail price equations		
Beef and veal	-3.61 (1)	-5.73** (1)
Pork	-3.18 (3)	-5.74** (3)
Poultry	-4.65 (0)	-4.44 (1)
Eggs	-3.94 (4)	-3.80 (1)
Dairy	-4.84 (1)	-7.52** (1)
Fresh fruit	-4.20 (0)	-4.12 (1)
Fresh vegetables	-4.22 (2)	-4.55 (2)
Farm prices (ln P_f)		
Beef and veal	-3.18 (0)	-3.15 (1)
Pork	-3.41 (0)	-3.49 (1)
Poultry	-3.41 (1)	-5.21* 1)
Eggs	-3.15 (1)	-4.99 (1)
Dairy	-3.72 (1)	-6.00** (1)
Fresh fruit	-4.02 (0)	-3.92 (1)
Fresh vegetables	-3.82 (1)	-5.23* (1)

Values are t-tests associated with the coefficient of the lagged OLS residuals in which the regression includes a constant and a time trend. Values in parentheses indicate the number of lagged-first differences in the autoregression (Dickey-Fuller) or the number of lags included in the error covariogram (Phillips-Perron).

*Reject the null of a spurious regression at (approximately) the 0.10 level. The result is based on a critical value of approximately -5.2. which is -0.5 plus -4.7. -4.7 is the critical value computed by Phillips and Ouliaris for a demeaned (constant) and detrended (one deterministic time trend) regression with five stochastic regressors (Phillips and Ouliaris, Table IIc). -0.5 is the increment associated with adding the sixth stochastic regressor (Ng).

**Reject the null of a spurious regression at (approximately) the 0.05 level. The result is based on a critical value of approximately -5.5 which is -0.5 plus -5.0. -5.0 is the critical value associated with a demeaned and detrended regression with five stochastic regressors and -0.5 is the increment associated with the sixth regressor (Ng).

To employ the test, however, one is faced with choosing a set of integrated and potentially superfluous variables to append to the model. Here, theory provides no clear guide. Hence, we follow Park's suggestion by appending higher order polynomials of deterministic time trends to the retail and farm price equations.

Table 2 reports chi-square estimates associated with Park's J_1 test (1990) of the null that the coefficients of the polynomial time trend terms (T^2, T^3 & T^4) are jointly zero (see table 2 for the specification of the polynomial terms). The p-values (in parentheses) suggest that at the 0.05 level of rejection, all equations are stochastically cointegrated except the retail price equation for poultry and the farm price equations for beef, poultry, and dairy.

The results in tables 1 and 2 provide somewhat mixed results. The Dickey-Fuller tests (table 2) suggest each of the 14 equations is spurious. The Phillips-Perron tests suggest that, at reasonable levels of rejection, 8 of the 14 price equations are cointegrated. Finally, Park's J_1 test suggests that, at reasonable levels of rejection, 10 of the 14 equations are cointegrated. In combination, the Phillips-Perron and Park's test results suggest that only the retail price equation for poultry and the farm price equation for beef and veal may be spurious.

Table 2 — Variable-addition tests of cointegration

Industry	Retail price equations	Farm price equations
Beef and veal	6.83 (.08)	9.52 (.02)
Pork	3.75 (.29)	0.59 (.90)
Poultry	18.23 (.00)	27.27 (.00)
Eggs	4.81 (.19)	3.16 (.37)
Dairy	7.72 (.05)	16.90 (.00)
Fresh fruit	1.28 (.73)	6.83 (.08)
Fresh vegetables	0.31 (.96)	2.92 (.40)

Entries are c2 with the restriction $a1=a2=a3=0$ in the general regression $P=\mathbf{X\beta} + \mathbf{f(a,T)} + e$ in which $\mathbf{f(a,T)} = {}_1T^2 + a_2T^3 + a_3T^4$ and in which T = t/max(t) where P is either the logged and deflated retail or farm price, X includes the intercept, the linear time trend, and the six logged and deflated stochastic regressors. The values in parentheses are p-values, and represent the size of the rejection region associated with the restriction. For example, values greater than 0.05 fail to reject the null of stochastic cointegration at the 0.05 level of rejection.