

## **Assessment of RSTEM Forecasts**

**Andrew J. Buck  
Economics Department  
Temple University  
215-204-1985  
buck@temple.edu**

### **Abstract**

This paper reports on forecast analyses conducted with the Regional Short Term Energy Model (RSTEM). Forecasts were developed for consuming sectors (residential, commercial and industrial) using the October 2007 release of RSTEM. The forecasted fuels were natural gas, electricity, and distillate. For natural gas and electricity the forecasts were generated for the nine census regions. The commercial and industrial distillate forecasts for geographic regions were based on the set of fourteen areas created from the nine census regions. Residential distillate consumption forecasts were developed for four geographic aggregates of the nine census regions. For each fuel -- region doublet the October forecast was compared to the realized data reported in May 2008, using % error. The October model was rerun using the realized values for the exogenous variables and the endogenous variable add factors. The % forecast error was again calculated. Finally, the October model was rerun using the realized values for the exogenous variables and no add factors for the endogenous variables. The % forecast error was again calculated.

# Assessment of RSTEM Forecasts<sup>1</sup>

**Andrew J. Buck**  
**Economics Department**  
**Temple University**  
**215-204-1985**  
**buck@temple.edu**

**I. Background.** This paper is part of a series of reports presented to the ASA Committee on a project to develop methods for evaluating EIA forecasts of energy production, consumption, and other projected attributes of the U.S energy system. Prior reports concentrated on projections made for the *Annual Energy Outlook(AEO)* based upon the National Energy Modeling System (NEMS).

The results presented here examine RSTEM forecast errors for consumption of natural gas, electricity, and distillate for the commercial, industrial and residential sectors. The geographic disaggregation corresponds to that used in RSTEM. For natural gas and electricity the nine census regions were used. For commercial and industrial distillate consumption there are fourteen geographic regions made up from the nine census regions. Residential distillate is modeled for only four geographic regions in RSTEM. For each fuel-sector-region triplet a set of three forecast errors was computed for the period October 2007 through March 2008. The three forecast errors are the out-of-sample error, the error made when using actual exogenous variable data for the forecast period with endogenous variable add factors, and the error made when using actual exogenous variable data but not add factors.

**II. Analysis of Differences between Actual and Forecast Values For Selected RSTEM Demand Projections.** A goal of the forecast evaluation project is to measure the differences between projections and the eventual outcomes of the projected series. Once done, the analysis is additionally to identify those portions of the difference found that can be attributed to differences between the assumed and actual values of the explanatory variables on which the projected series depends. The method used has been to compare the out-of-sample forecast with the forecast made after the exogenous variables have become known and applying the RSTEM staff's manually implemented add factors. The comparison is also made while ignoring the add factors. The latter comparison provides a way to evaluate the accuracy of the judgmental component of RSTEM.

**II.1 Method.** The October 2007 release of RSTEM was the basis for the results presented here. The model release changes from month to month as equations are re-estimated and sectors of the model are re-specified. The model's exogenous variables include a variety of weather,

---

<sup>1</sup> This is a working document prepared for the Energy Information Administration (EIA) in order to solicit advice and comment on statistical matters from the American Statistical Association Committee on Energy Statistics. This topic will be discussed at EIA's fall 2008 meeting with the Committee to be held October 24, 2008

demographic and economic data obtained from a number of US Government agencies. In addition, the RSTEM staff include add factors that they assign judgmentally. An add factor is a small number added to the intercept of a behavioral equation in the system with the intent of improving the accuracy of the forecast of the endogenous variable. There were 407 add factors applied in the October release of RSTEM. This is about 10% of all the entire model's endogenous variables. Hence, the accuracy of the RSTEM forecast depends on the validity of the model, the accuracy of the exogenous variable forecasts obtained from other agencies, and the judgment of the RSTEM staff in applying add factors.

**Percent Error** - The endogenous variable forecasts under the three applications of the October model were compared to the realized data for the period October 2007 through March 2008 as reported by the June 2008 release of the Short Term Energy Outlook. The method of comparison was to compute the month by month percent error of the forecast. Algebraically this amounted to

$$\% \text{ error} = (\hat{y}(j,k)_t - y_t) / y_t$$

Where  $\hat{y}$  is the forecast, j is either unknown or known exogenous data for forecast period, k is either add factor or no add factor, and y is the realized value of the endogenous variable.

**Naming** - In the following material the naming convention is based on the mnemonics used by RSTEM. As a matter of convention the variable names in the graphs that follow are preceded by PCT to denote percent error, per the above formula. After the under score, the first two characters connote the fuel, the next two are used to connote the consuming sector, and the last character, always P or X, connotes physical units. After the second underscore the first three characters connote the geographic region. For distillate the geographic code becomes a two digit code instead of three letters. The last character in the name is always a digit. A 1 indicates an out of sample forecast error, a 2 denotes a forecast error using known exogenous variables and add factors, and a 3 denotes a forecast error using known exogenous variables but no add factors. For the purposes of the graphs, the convention is summarized in Table II.1

Fuels	Natural Gas	NG
	Electricity	EX
	Distillate	DK, D2
Consuming Sector	Residential	RC
	Industrial	IN or IC or CGCON
	Commercial	CC or CECON
Geographic Regions	ENC (03)	East North Central
	ESC (06)	East South Central
	MAC (02)	Middle Atlantic (less NY)
	MTN (08)	Mountain
	NEC (01)	New England
	PAC (09)	Pacific (less CA)
	SAC (05)	South Atlantic (less FL)
	WNC (04)	West North Central

	WSC (07)	West South Central (less TX)
	10	CA
	11	Florida
	12	NY
	13	TX
	14	Alaska & Hawaii

**Fuels and Regions** - The geographic regions and respective fuels are provided in Table II.2.

Table II.2	
Fuel	Regions
NGRCP, NGCCP, NGINX (Natural Gas) EXRCP, EXCCP, EXICP (Electricity)	NEC, MAC, ENC, WNC, SAC, ESC, WSC, MTN, PAC
D2RCX (Distillate, residential)	NE, SO, MW, WE
DKCECON (Distillate, commercial)	NEC, MAC less NY, ENC, WNC, SAC less FL, CA, TX, NY, AK+HI (missing ESC, WSC less TX, MTN, PAC less CA)
DKCGCON (Distillate, industrial)	NEC, MAC less NY, ENC, WNC, SAC less FL, ESC, MTN, NY, AK+HI (missing WSC less TX, PAC less CA, CA, FL, TX)

The distillate geographic regions differ across consuming sectors due to zeros, as distinct from missing observations, in the data. The reason for the zeros is unclear. For example, in the DKCGCON series PAC less CA has a single non-zero entry, while WSC less TX has long strings of zeros with the non-zero entries varying by several orders of magnitude.

**Equation Specifications** - As a representative behavioral equation the residential demand for natural gas in the East North Central region is given by:

$$\begin{aligned}
 & (\text{NGRCP\_ENC}) / ((\text{QHNGRECS\_ENC} / \text{QHALLRECS\_ENC}) * \text{QHALLC\_ENC}) \\
 & = 0.000696490213337 \\
 & - 2.47001192173e-05 * \text{FEB} - 0.000182046614417 * \text{MAR} - 0.000341310981 * \text{APR} - \\
 & 0.000491956117136 * \text{MAY} - 0.000566242675272 * \text{JUN} - 0.000593361051001 * \text{JUL} - \\
 & 0.000598536863863 * \text{AUG} - 0.000575695941079 * \text{SEP} - 0.000484885164602 * \text{OCT} - \\
 & 0.000310944850903 * \text{NOV} - 0.000123400946201 * \text{DEC} \\
 & + 1.23621611286e-05 * \text{HR\_D\_ENC} \\
 & + 4.53082156325e-06 * \text{HR\_D\_ENC}(-1) \\
 & - 0.75433467033 * \text{RELCOSTRG\_ENC}(-1) \\
 & - 7.12026425869e-08 * \text{TIME}
 \end{aligned}$$

Residential natural gas consumption is scaled by the proportion of households with natural gas hookups that are actually using them. The variable HR\_D\_\* and RELCOSTRG are constructed variables defined as heating requirements and relative costs of gas, respectively. The price of natural gas enters the equation through RELCOSTRG, and is exogenous since it is lagged one period.

The add factor mentioned above is a series of small numbers added to the intercept so that the equation fits the data exactly. In sample these numbers are actually the least squares residuals. For out-of-sample forecasts the RSTEM staff specifies the values judgmentally in order to give a better forecast. Part of the present exercise provides an assessment of whether their judgment is correct.

The specifications for the demand equations vary across fuels and regions, but can be summarized in Table II.3:

Table II.3 Equation Specifications		
Fuel and Sector	Left Hand Side	Right Hand Side*
Natural Gas: Residential	Natural Gas consumption per consuming household	- Monthly dummies - HDD deviation from normal weighted by # of NG hookups - Lagged real expenditure on NG weighted by real wages <sup>1</sup>
Natural Gas: Industrial	Industrial Natural Gas Consumption as Shares of Total Industrial Consumption <sup>2</sup>	-Monthly dummies -HDD deviation from normal weighted by total industrial NG consumption - Industrial production index - PDL <sup>3</sup> on real expenditure on NG
Natural Gas: Commercial	Commercial NG Consumption, unscaled	-Monthly dummies - Private sector employment - HDD deviation from normal - PDL on lagged NG expenditure <sup>1</sup>
Electricity: Residential	Rate of growth of consumption per household	- Lagged rate of growth of consumption - Lagged real price or price growth - Lagged expenditure - Lags on HDD and CDD - Event dummies
Electricity: Industrial	Log of consumption or rate of growth	- Lag of dependent variable - Lag of real expenditure on electricity <sup>1</sup> scaled by real wage disbursements - Employment - Event dummies
Electricity: Commercial	Log of consumption or rate of growth	- Lagged rate of growth - Change in CDD - Change in HDD - Rate of growth of employment - Expenditure on electricity <sup>1</sup> - Wage disbursements - CDD and HDD - Private sector employment
Distillate: Residential	Heating oil consumption weights	- Monthly dummies - Deviation of HDD from normal weighted by oil hookups

		- Event dummies
Distillate: Industrial <sup>4</sup>	Physical units	- Monthly dummies - Industrial production - CDD as deviation from normal - Price of NG - Event dummies
Distillate: Commercial <sup>4</sup>	Physical units	- Monthly dummies - Commercial employment - CDD as deviation from normal - Event dummies
<p>* For a given fuel and sector not all of the RHS variables are included in the specification for every geographic region.</p> <ol style="list-style-type: none"> <li>1. The price of the fuel enters here. Since the variable is lagged we treat price as predetermined and therefore exogenous.</li> <li>2. Actual variable inclusion and use of the PDL differs across regions. Across regions the equation is either log-log or semilog.</li> <li>3. PDL denotes Almon's polynomial distributed lag.</li> <li>4. This is actually distillate consumed for the purpose of generating electricity. The set of exogenous variables is found by racing back through various identities to the behavioral equations of the electricity model.</li> </ol>		

**II.2 Results.** Data for the forecast comparisons were taken from the October 2007 and June 2008 releases of the RSTEM databases. The following graphs present the results for each fuel: natural gas, electricity, and distillate. For each fuel the first three graphs show the out-of-sample forecast error, the forecast error when exogenous variables are known and add factors are used, and the forecast error when the exogenous variables are known by the add factors are omitted. The same three % errors are then presented as a triplet for each geographic region.

**Remarks:** *Residential Natural Gas* - Figures II.1 -- II.4 - The initial period out-of-sample forecast error is quite large for all the regions except West South Central. The Add Factors improve the forecasts for the six month ahead forecast. Looking at the nine regions individually, the add factors don't appear to be numerically important except at the 6 months ahead forecast.

*Industrial Natural Gas* - Figures II.5 -- II.8 - In the out-of-sample 2nd through 4th month ahead forecasts, most regions under predict consumption. The knowledge of the exogenous variables and the add factors cause the forecasts to improve except for West South Central. The graphs for each of the separate nine regions show that the add factors are essential to improving the forecasts, although they could be numerically more important.

*Commercial Natural Gas* - Figures II.9 -- II.12 - Perfect foresight regarding the exogenous variables and the add factors cut the % forecast error by about half. The add factors are numerically important only for the 6 months ahead forecast.

*Residential Electricity* - Figures II.13 -- II.16 - The electricity forecast errors are much smaller than those for natural gas. Inclusion of the actual exogenous data and the add factors improve the forecasts, especially six months out.

*Industrial Electricity* - Figures II.17 -- II.20 - The forecast errors are reduced by inclusion of actual exogenous data and the add factors. Except for WSC and NEC the add factors make quite a difference. The add factors appear to be different from zero only for the 5 and 6 month ahead forecasts.

*Commercial Electricity* - Figures II.21 -- II.24 - The actual exogenous data and add factors make the forecasts for WNC, WSC and ESC worse.

*Residential Distillate* - Figures II.25 -- II.28 - Residential distillate consumption (heating oil) is only a small part of U.S. energy use. RSTEM aggregates residential distillate into just four regions. There do not appear to be any add factors for these series. The forecasts are the best for the Northeast. Use of the actual exogenous variable improves all four forecasts, but not markedly.

*Industrial Distillate* - Figures II.29 -- II.32 - The add factors are not important to the improvement of the forecasts. The use of the add factors actually made the forecast for MAC less NY worse. As a group, these errors are the largest for any fuel studied.

*Commercial Distillate* - Figures II.33 -- II.36 - The add factors and exogenous data forecasts had very tight forecasts for 5 and months out. For MAC and SAC the exogenous data and the add factors made the forecasts worse.

Figure II.1 Natural Gas: Residential: Out-of-Sample

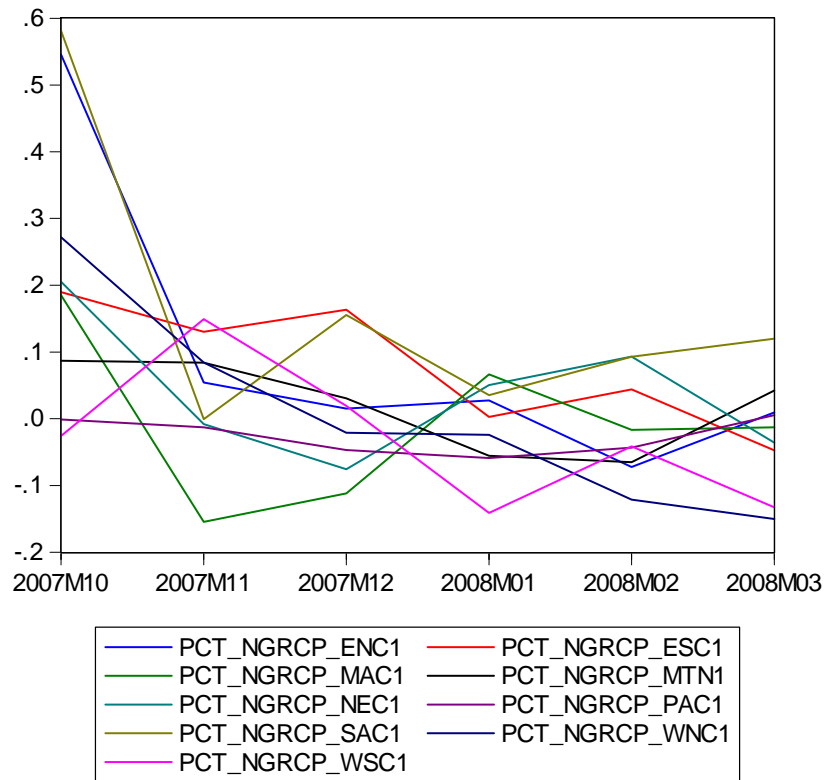


Figure II.2 Natural Gas: Residential: Exogenous Variables Known with Add Factors

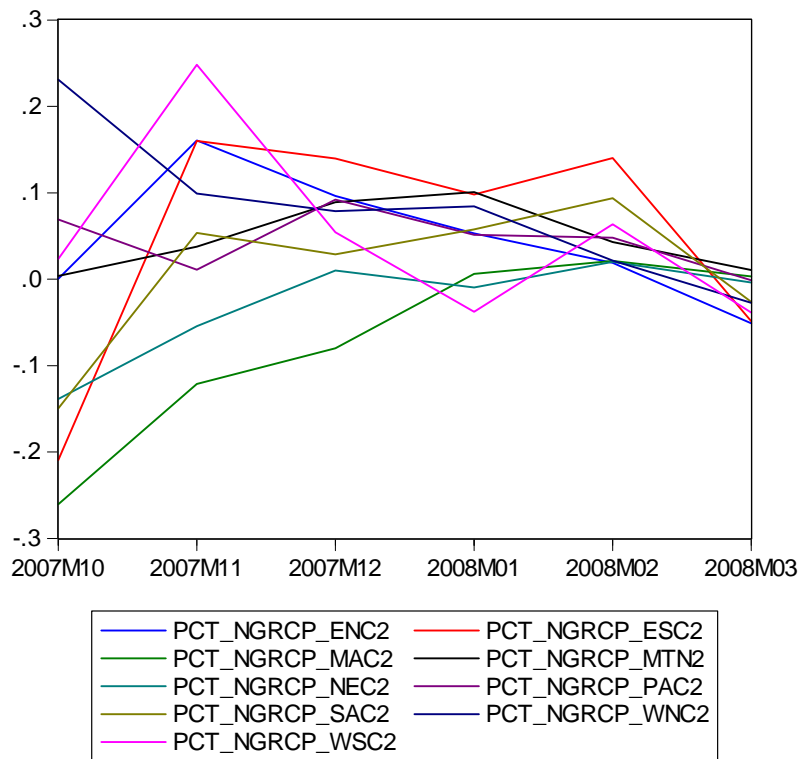


Figure II.3 Natural Gas: Residential: Exogenous Variables Known with No Add Factors

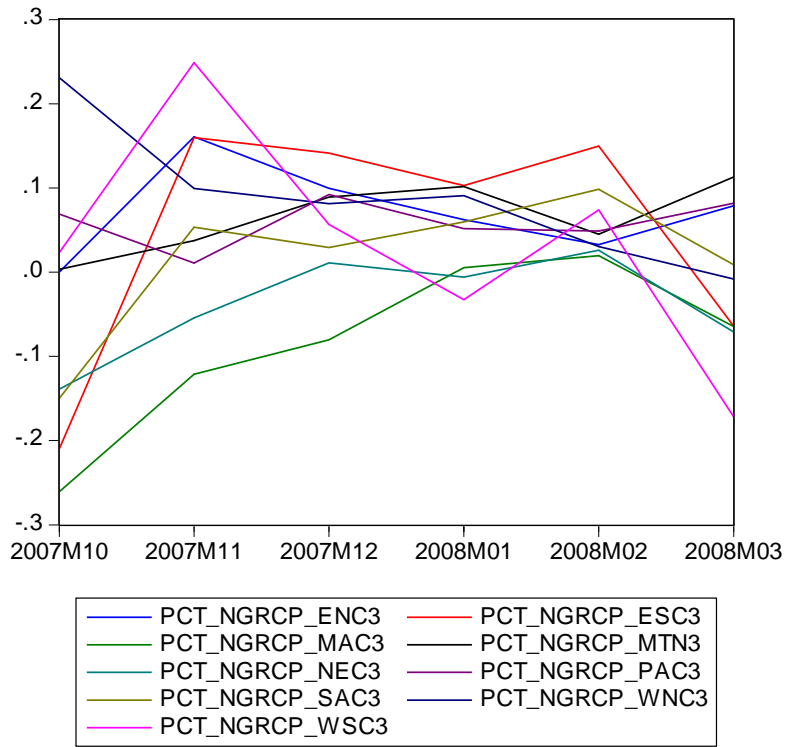


Figure II.4 Natural Gas: Residential: Blue=Out of Sample, Red=Exog w/ AF, Green=Exog w/out AF

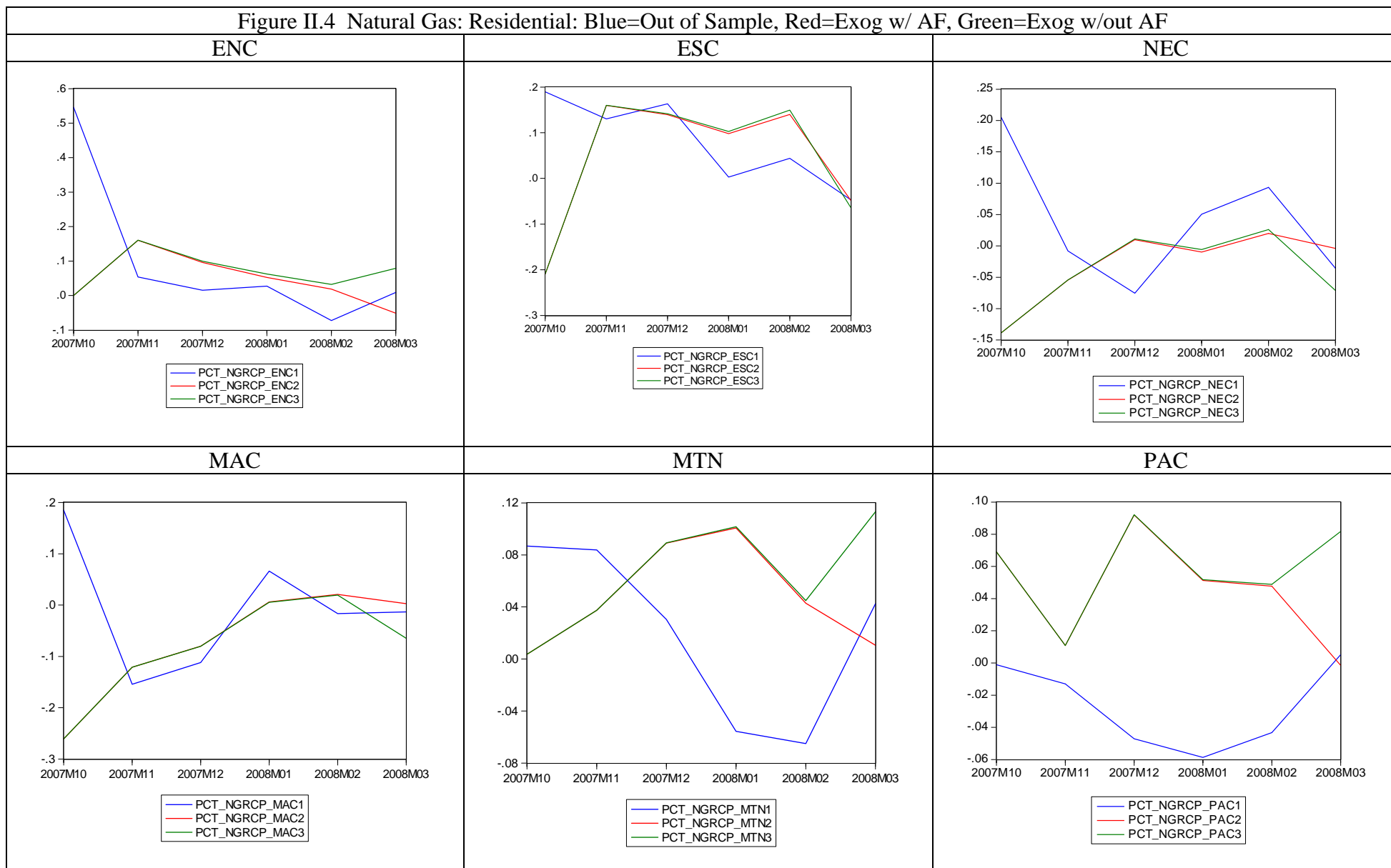


Figure II.4 (cont.) Natural Gas: Residential: Blue=Out of Sample, Red=Exog w/ AF, Green=Exog w/out AF

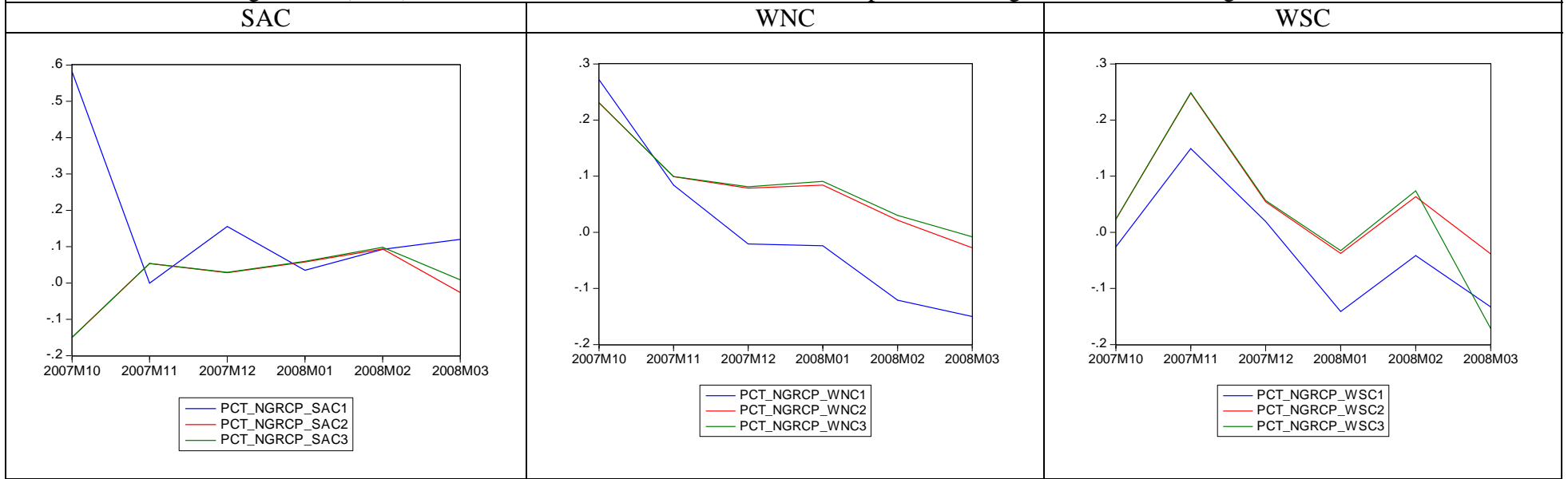


Figure II.5 Natural Gas: Industrial: Out of Sample

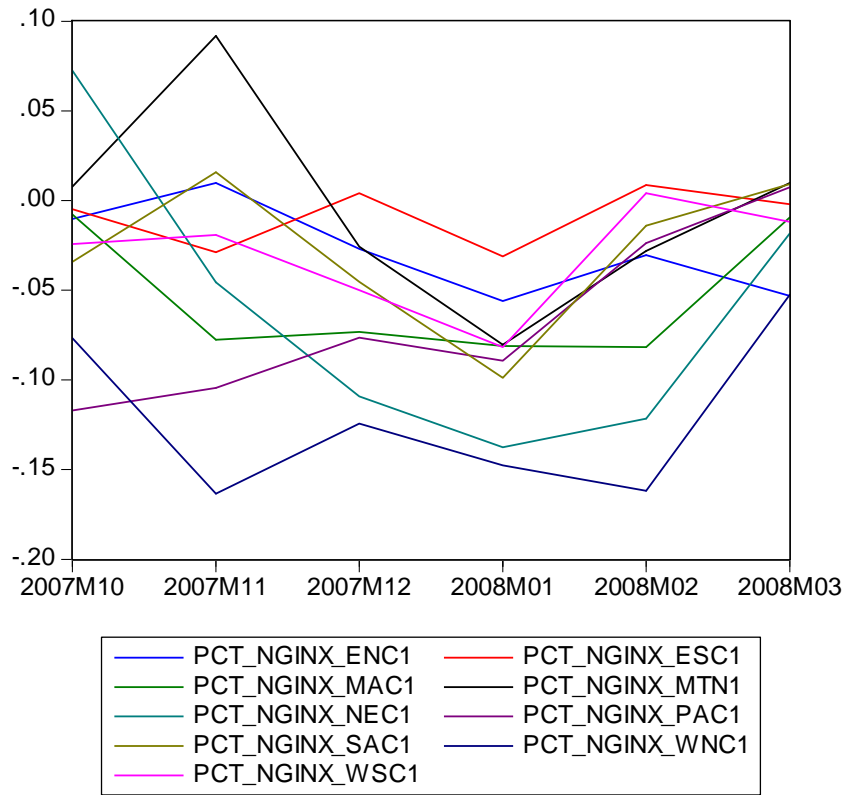


Figure II.6 Natural Gas: Industrial: Exogenous Known with Add Factors

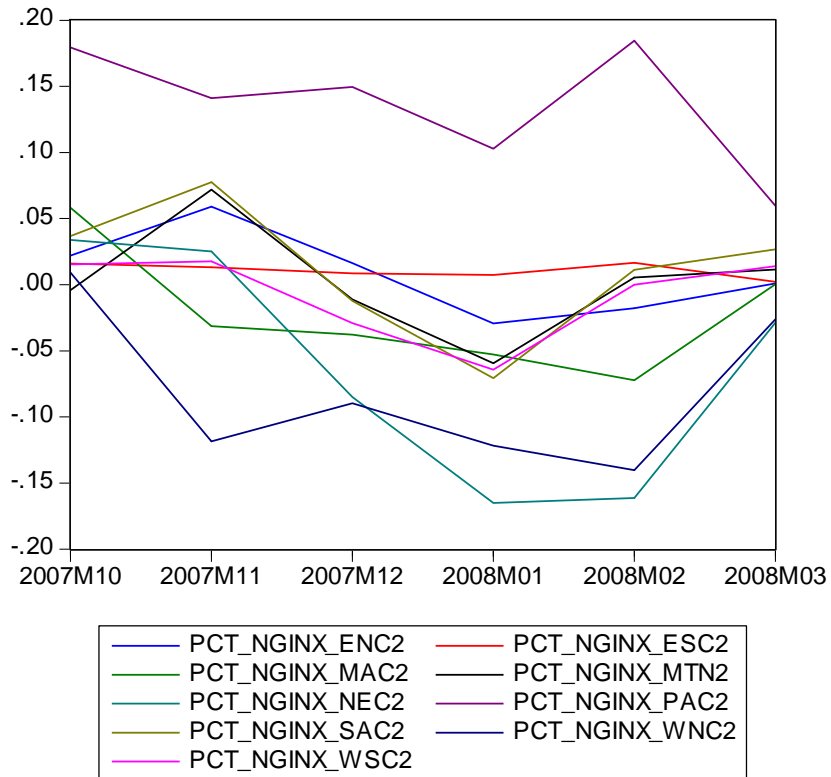


Figure II.7 Natural Gas: Industrial: Exogenous Known: No Add Factors

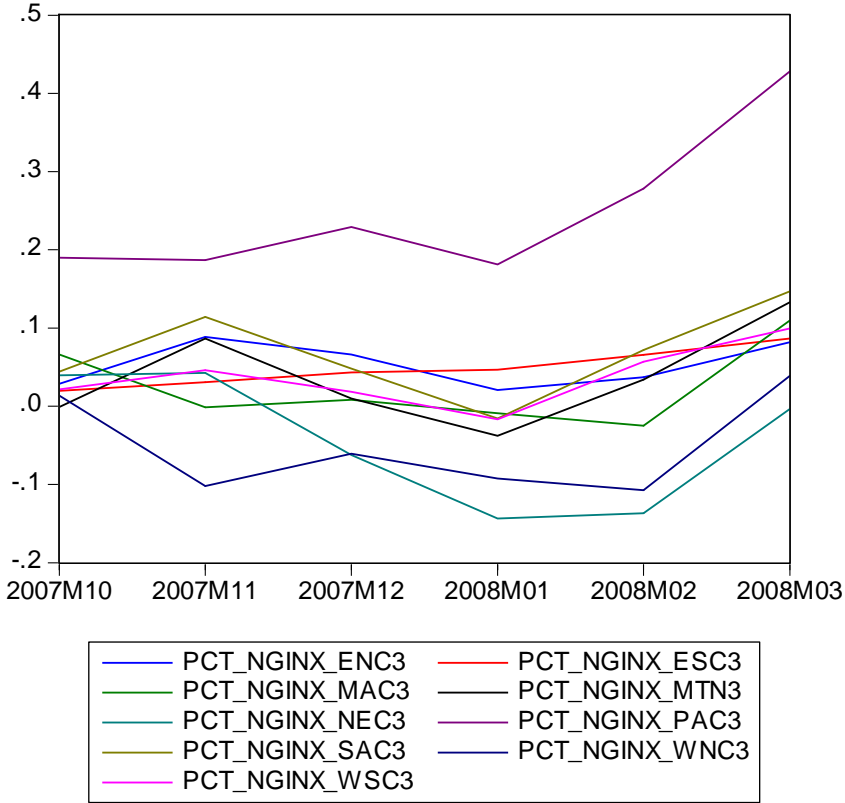


Figure II.8 Natural Gas: Industrial: Blue=Out of Sample, Red=Exog w/ AF, Green=Exog w/out AF

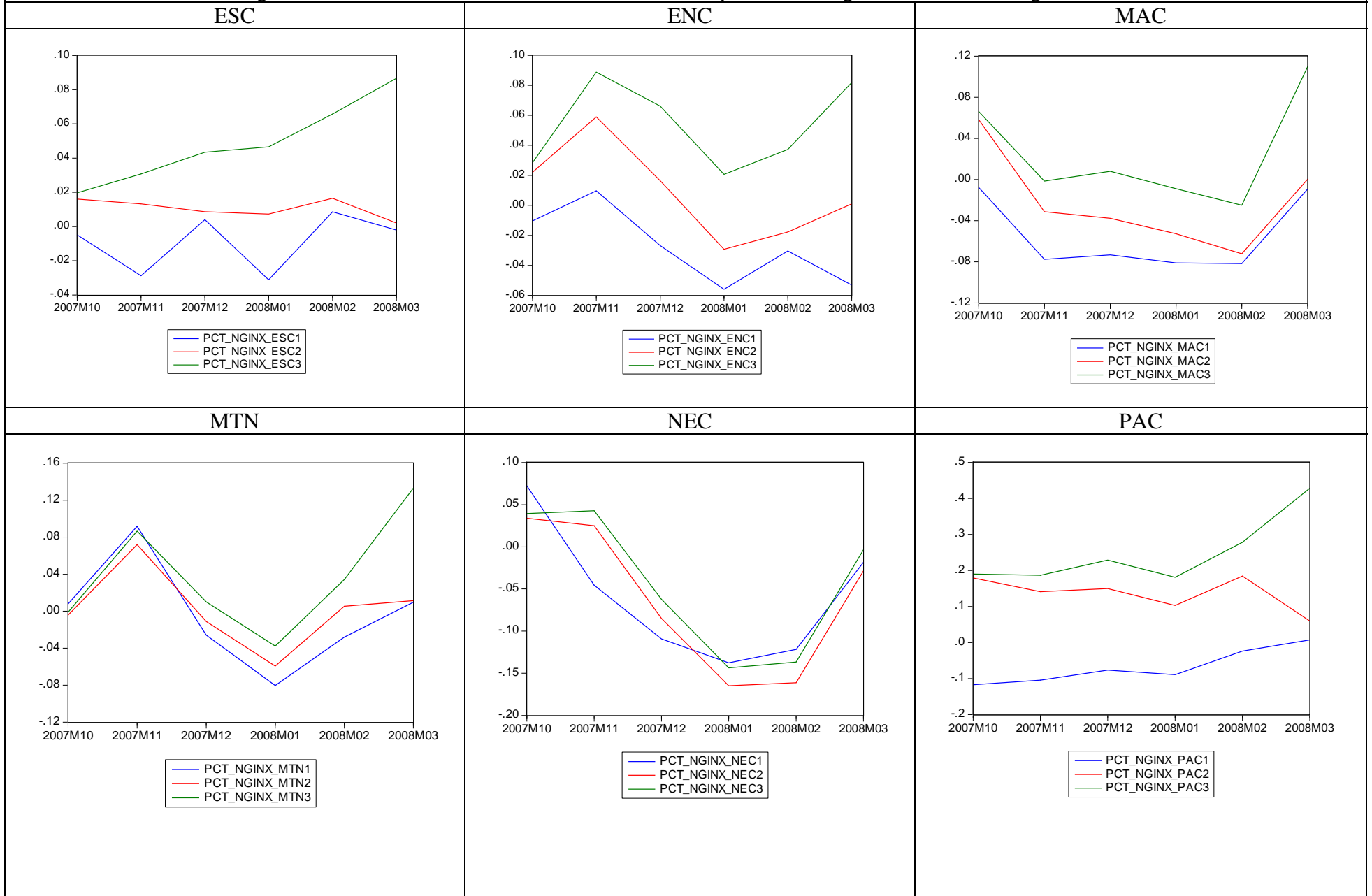


Figure II.8 (cont.)

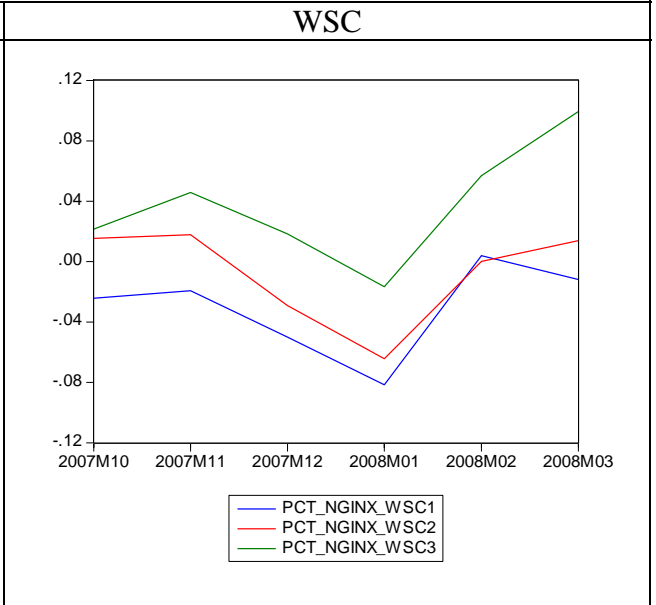
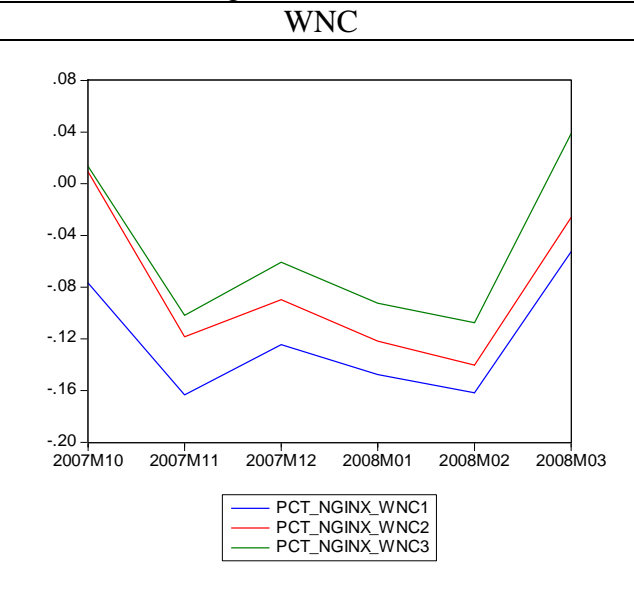
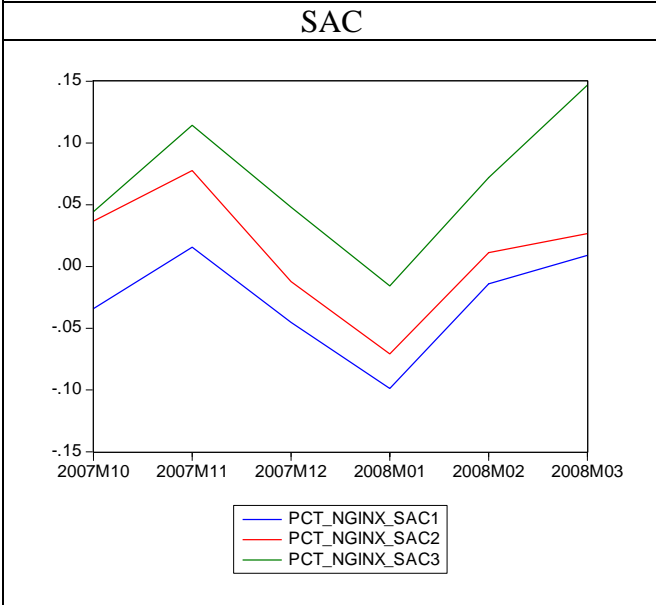


Figure II.9 Natural Gas: Commercial: Out of Sample

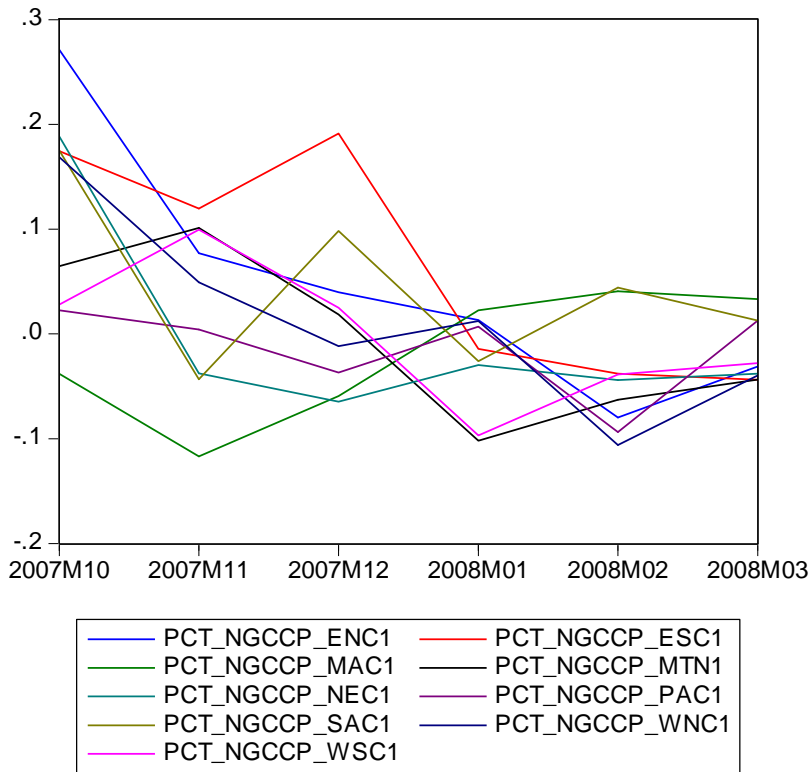


Figure II.10 Natural Gas: Commercial: Exogenous Known with Add Factors

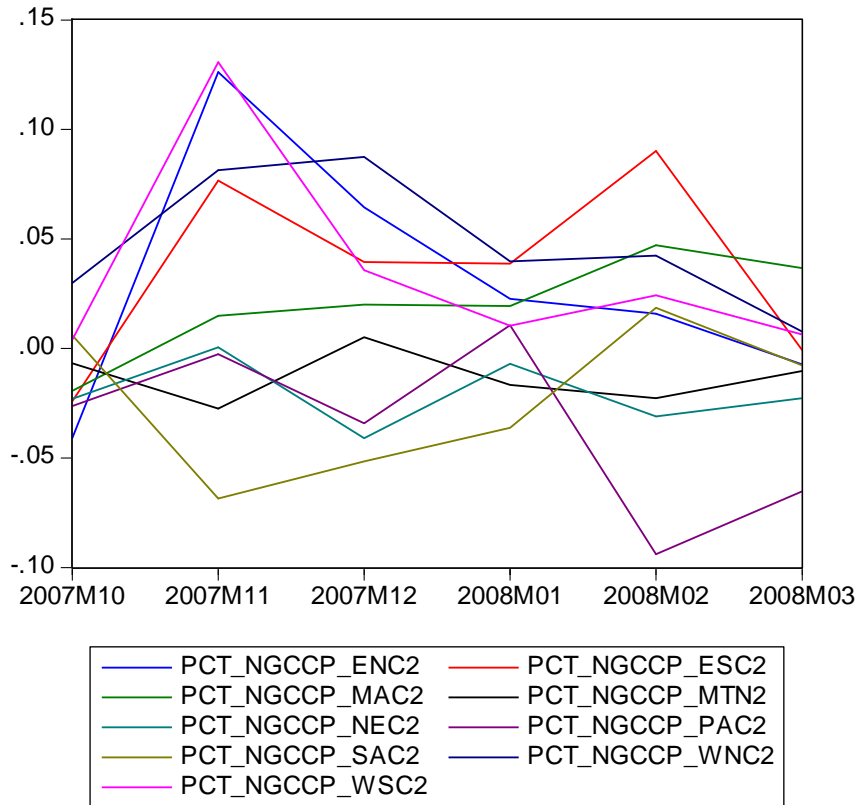


Figure II.11 Natural Gas: Commercial: Exogenous Known with No Add Factors

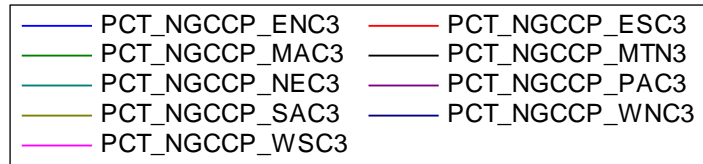
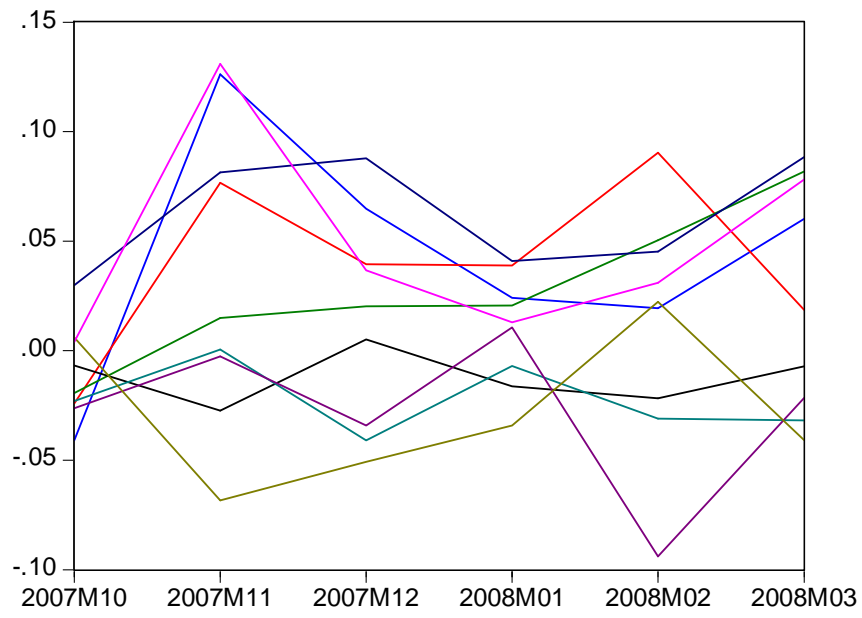


Figure II.12 Natural Gas: Commercial: Blue=Out of Sample, Red=Exog w/ AF, Green=Exog w/out AF

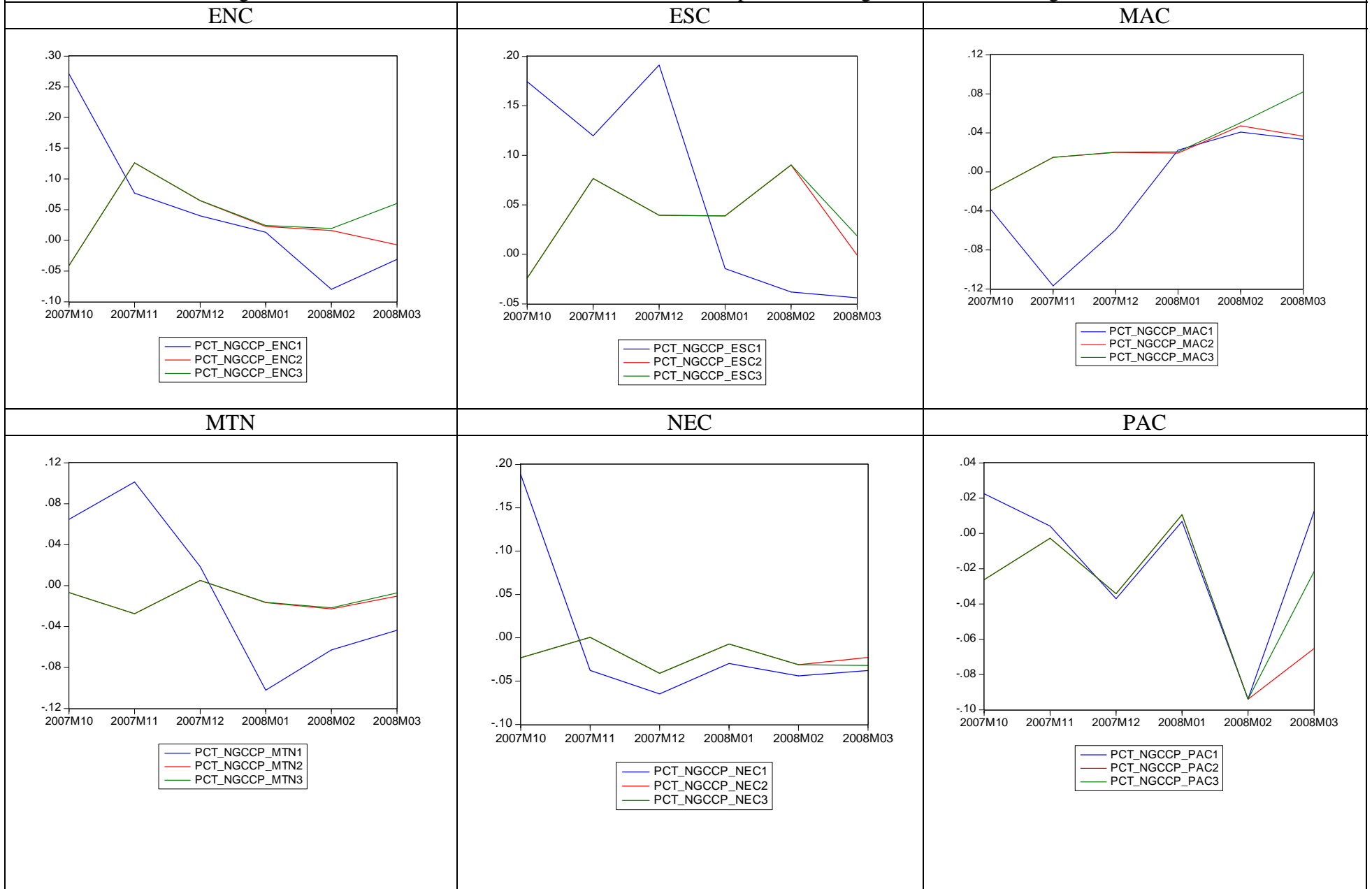
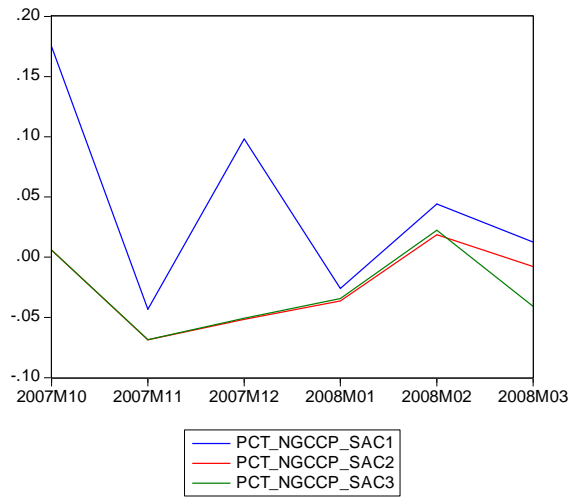
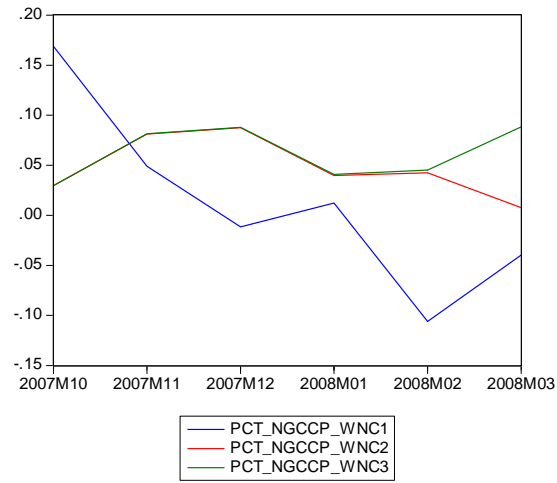


Figure II.12 (cont.)

SAC



WNC



WSC

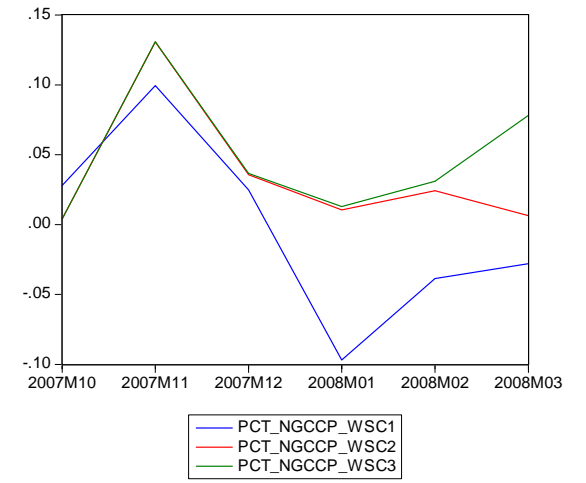


Figure II.13 Electricity: Residential: Out of Sample

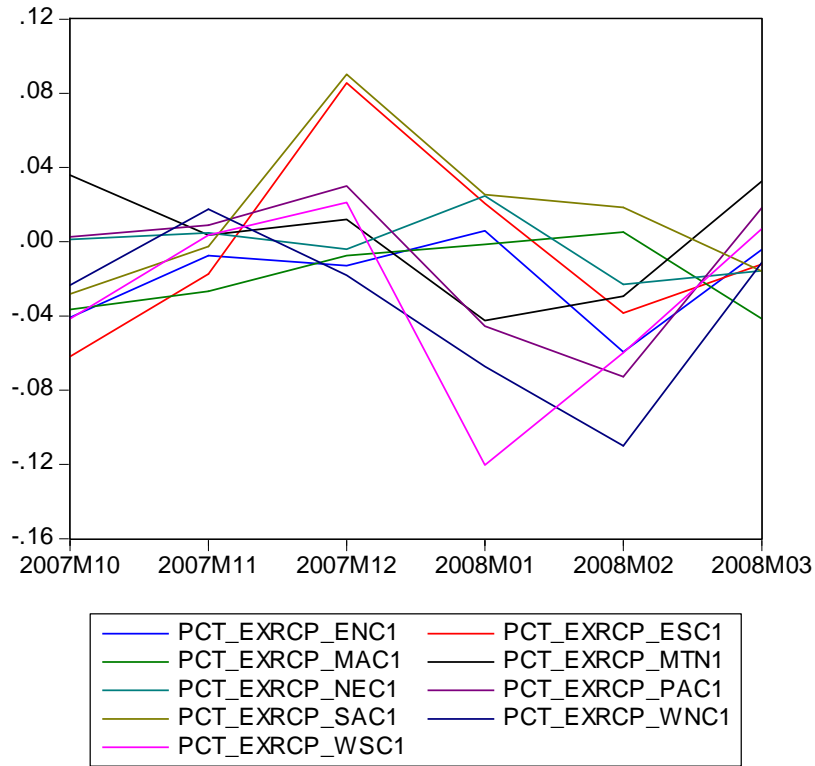


Figure II.14 Electricity: Residential: Exogenous Known with Add Factors

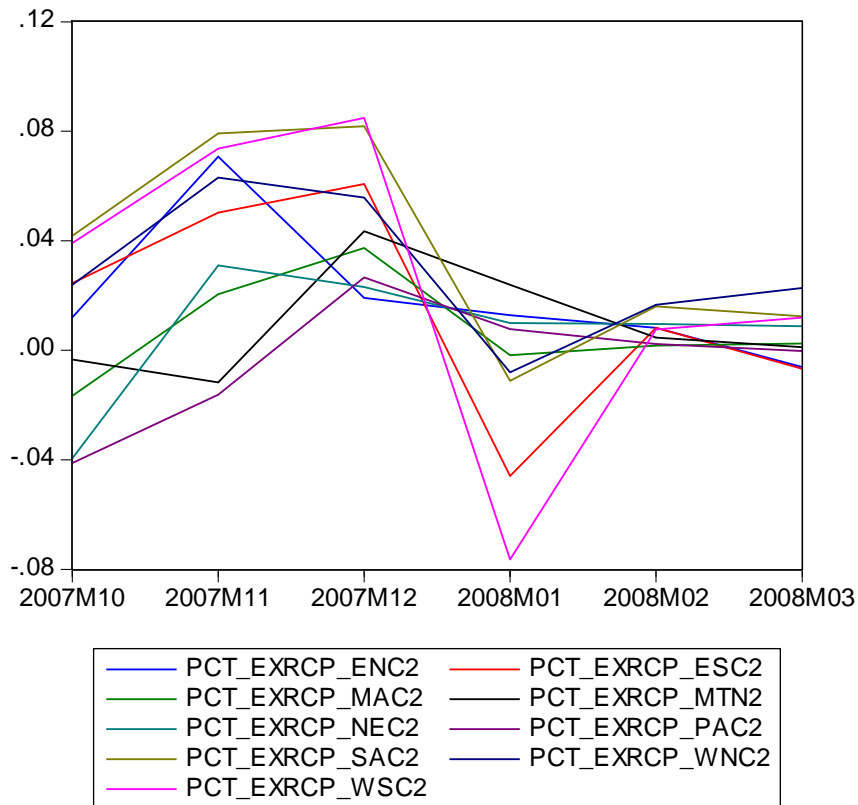


Figure II.15 Electricity: Residential: Exogenous Known with No Add Factors

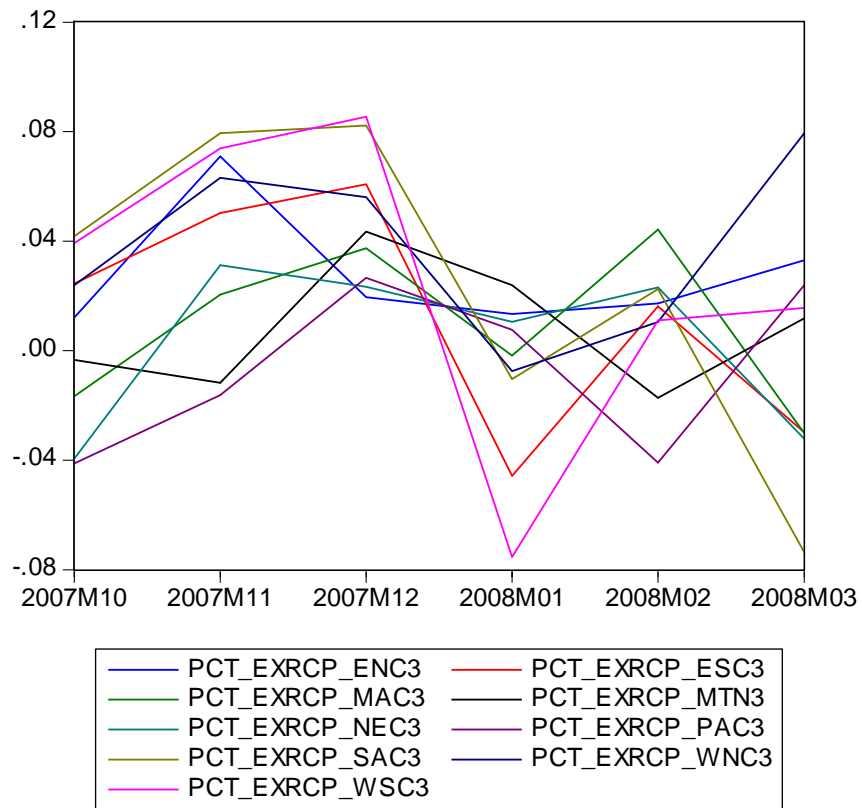


Figure II.16 Electricity: Residential: Blue=Out of Sample, Red=Exog w/ AF, Green=Exog w/out AF

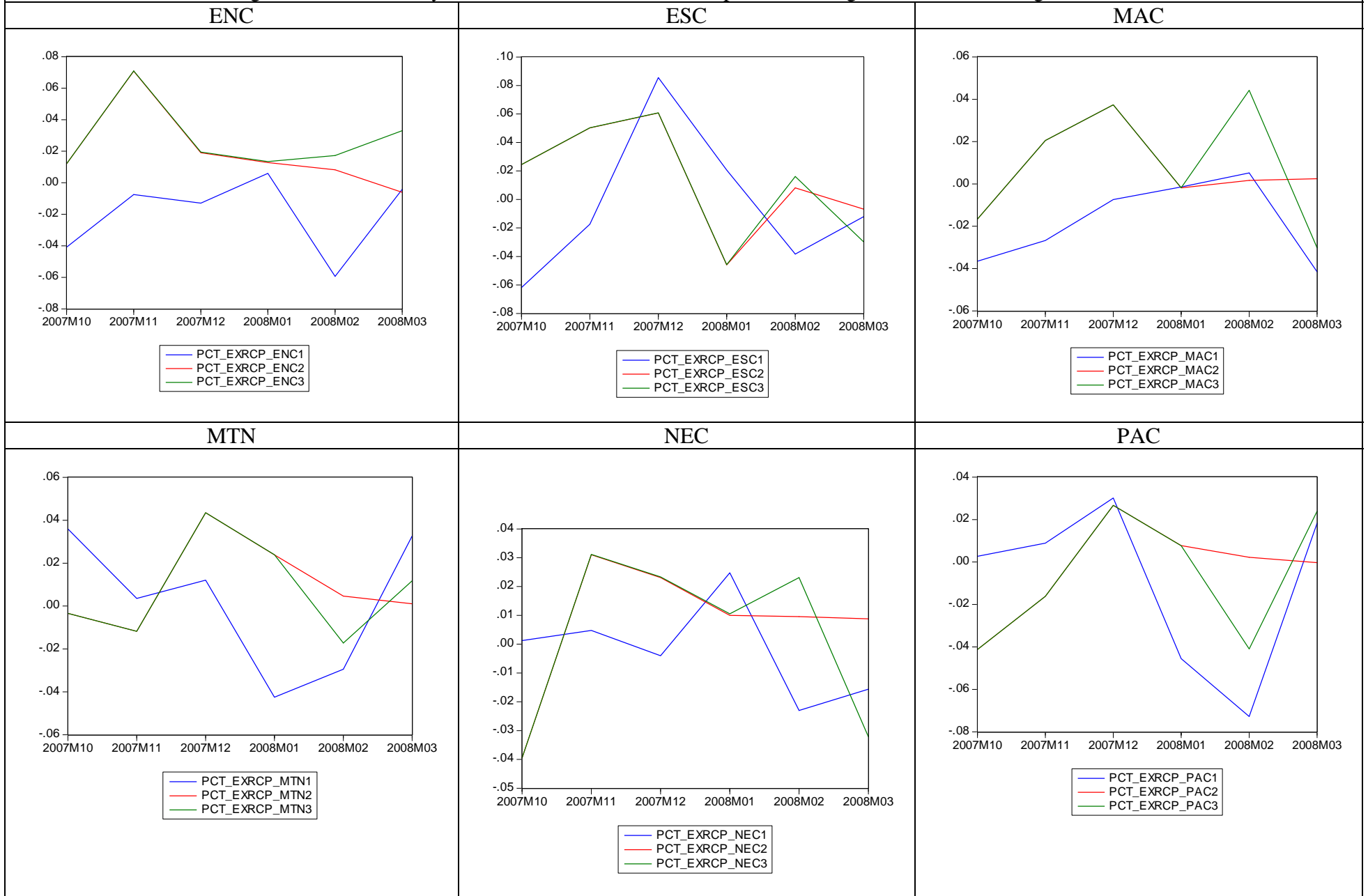
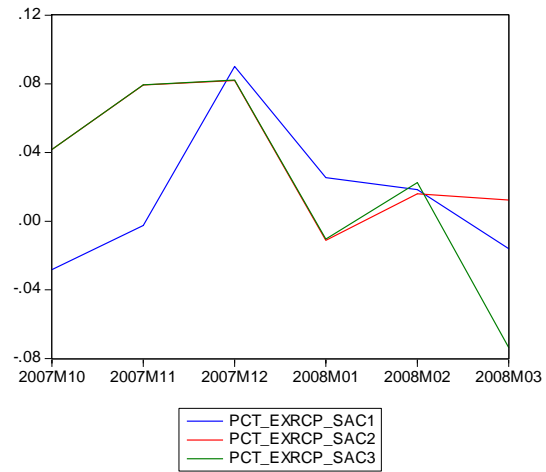
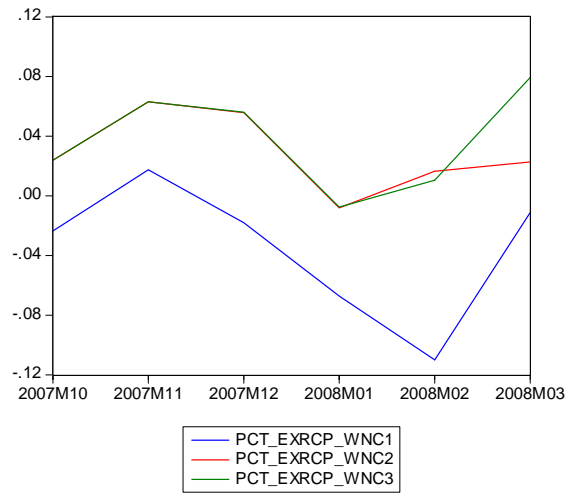


Figure II.16 (cont.)

SAC



WNC



WSC

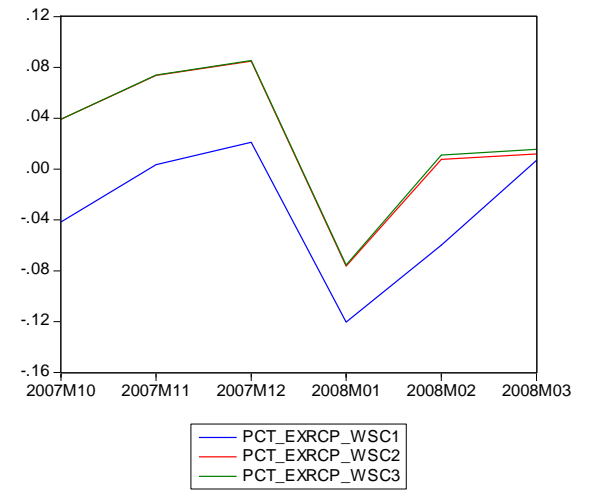


Figure II.17 Electricity: Industrial: Out of Sample

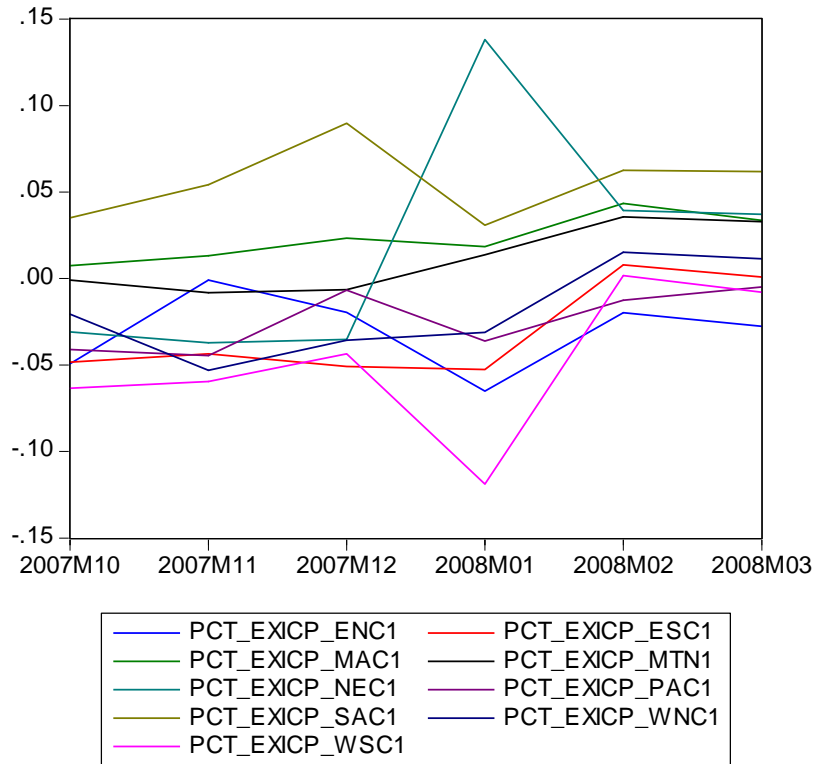


Figure II.18 Electricity: Industrial: Exogenous Variables Known with Add Factors

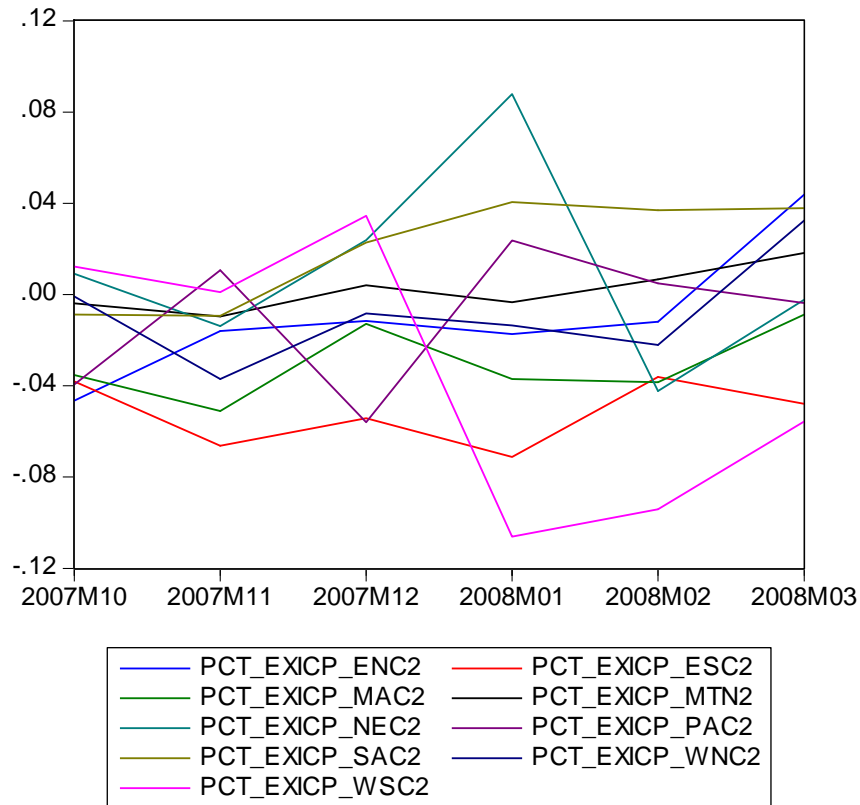


Figure II.19 Electricity: Industrial: Exogenous Variables Known with No Add Factors

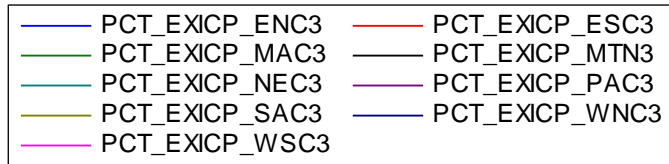
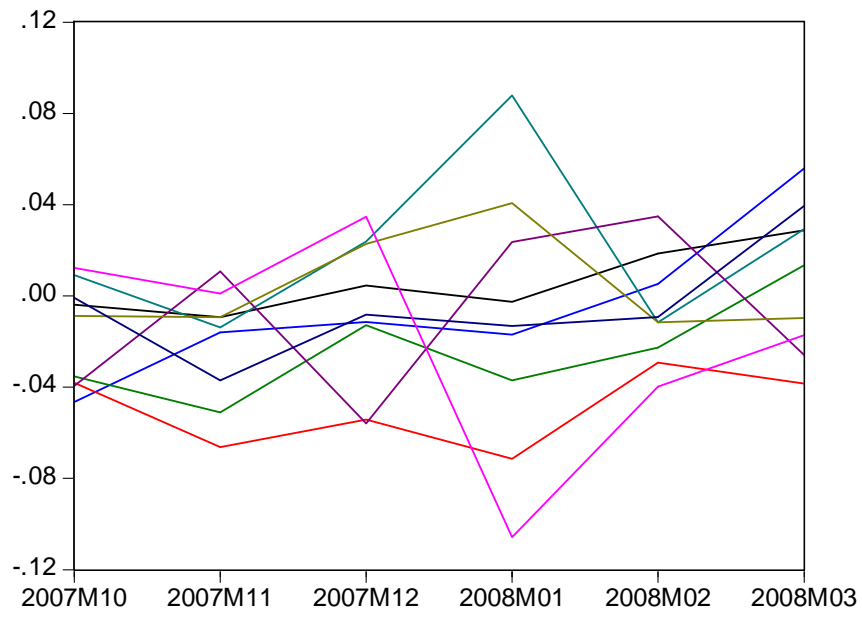


Figure II.20 Electricity: Industrial: Blue=Out of Sample, Red=Exog w/ AF, Green=Exog w/out AF

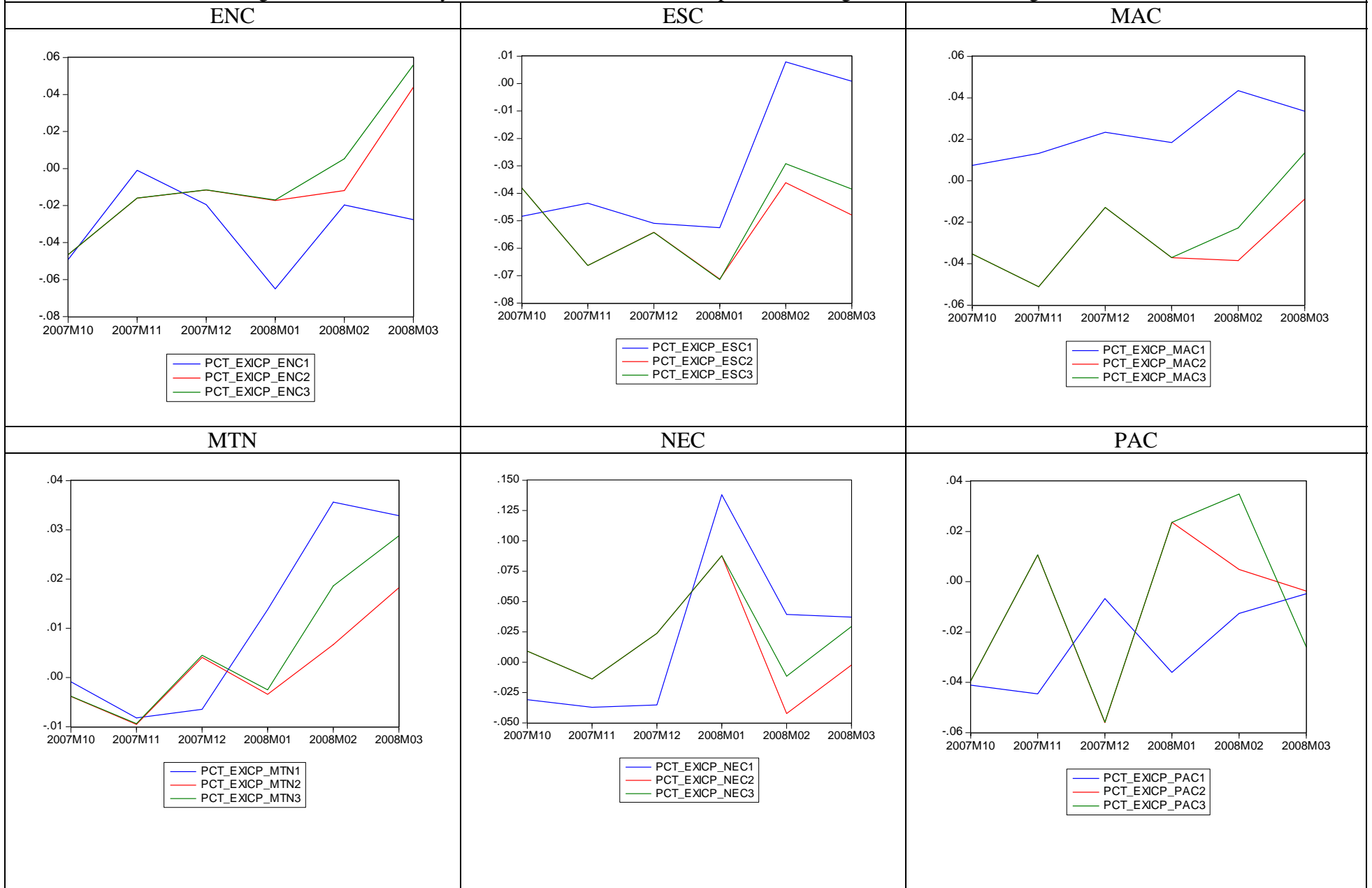
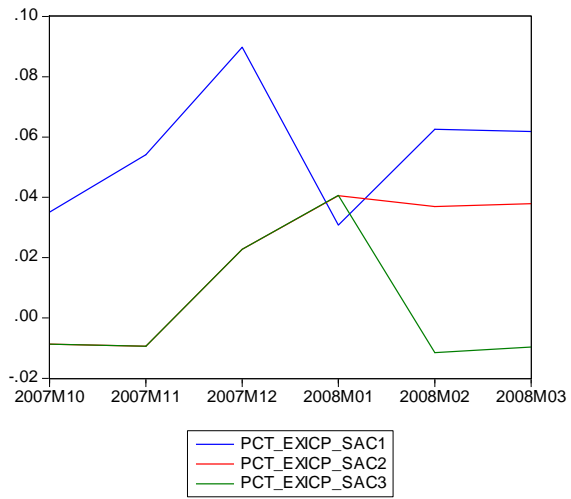
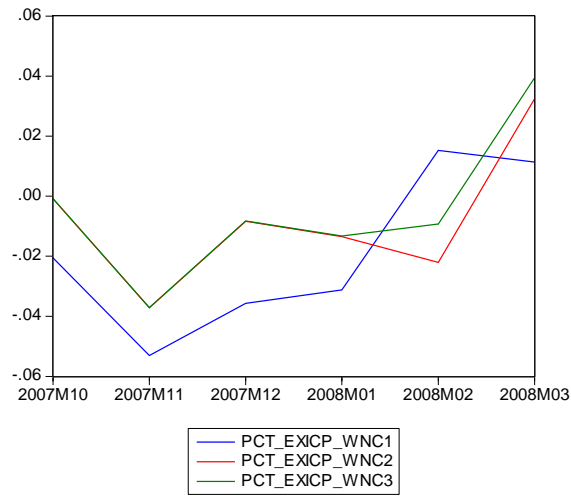


Figure II.20 (cont.)

SAC



WNC



WSC

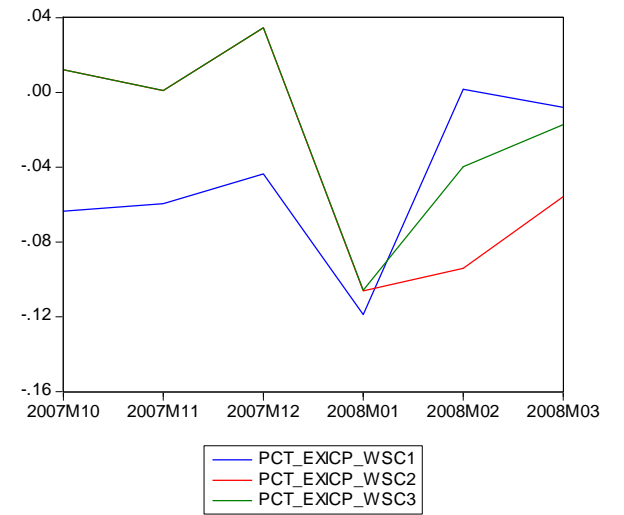


Figure II.21 Electricity: Commercial: Out-of-Sample

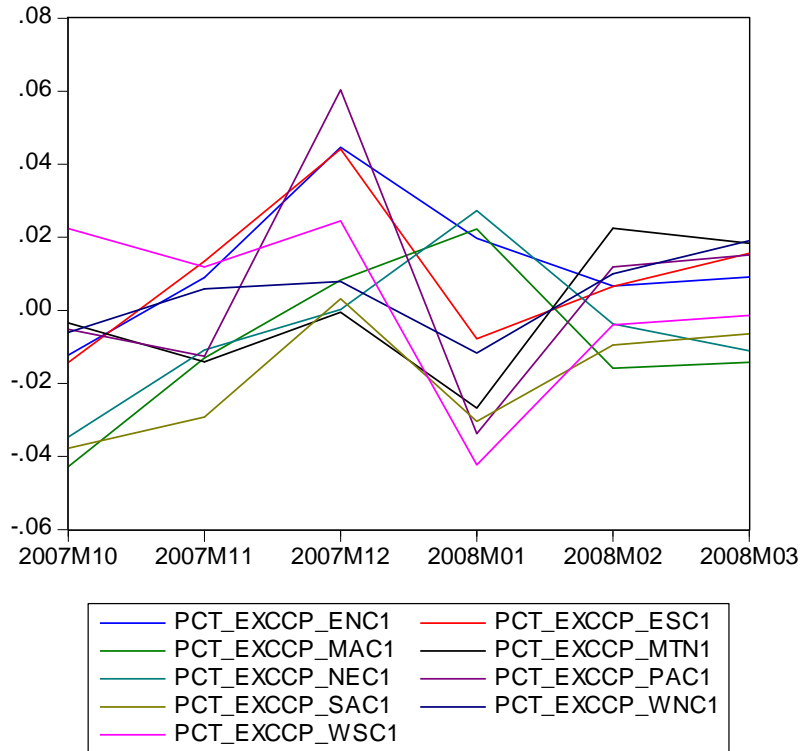


Figure II.22 Electricity: Commercial: Exogenous Variables Known with Add Factors

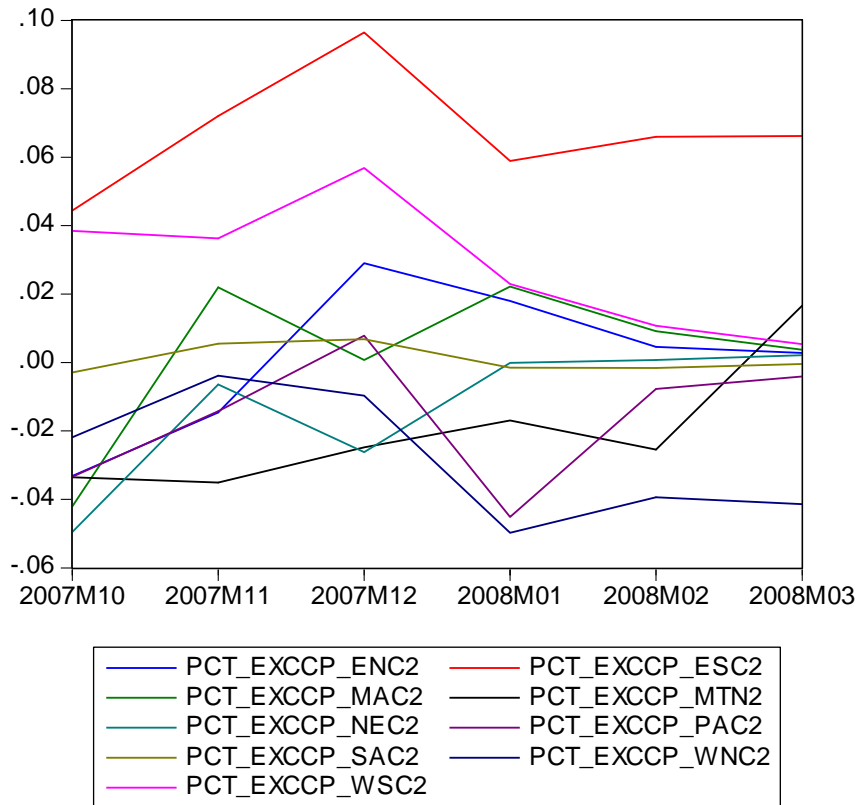


Figure II.23 Electricity: Commercial: Exogenous Variables Known with No Add Factors

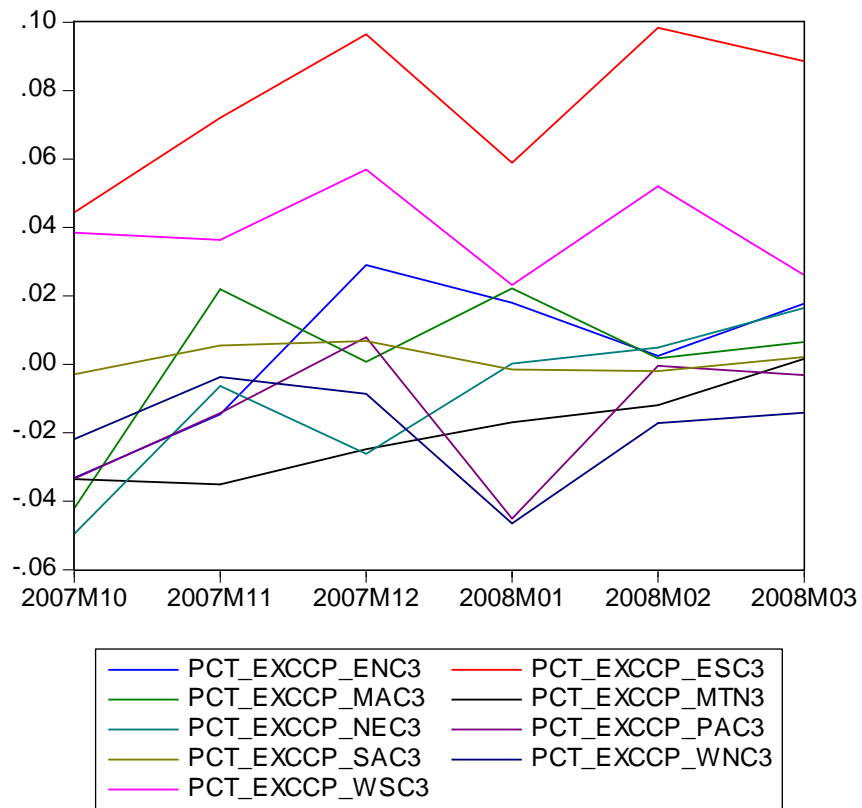


Figure II.24 Electricity: Commercial: Blue=Out of Sample, Red=Exog w/ AF, Green=Exog w/out AF

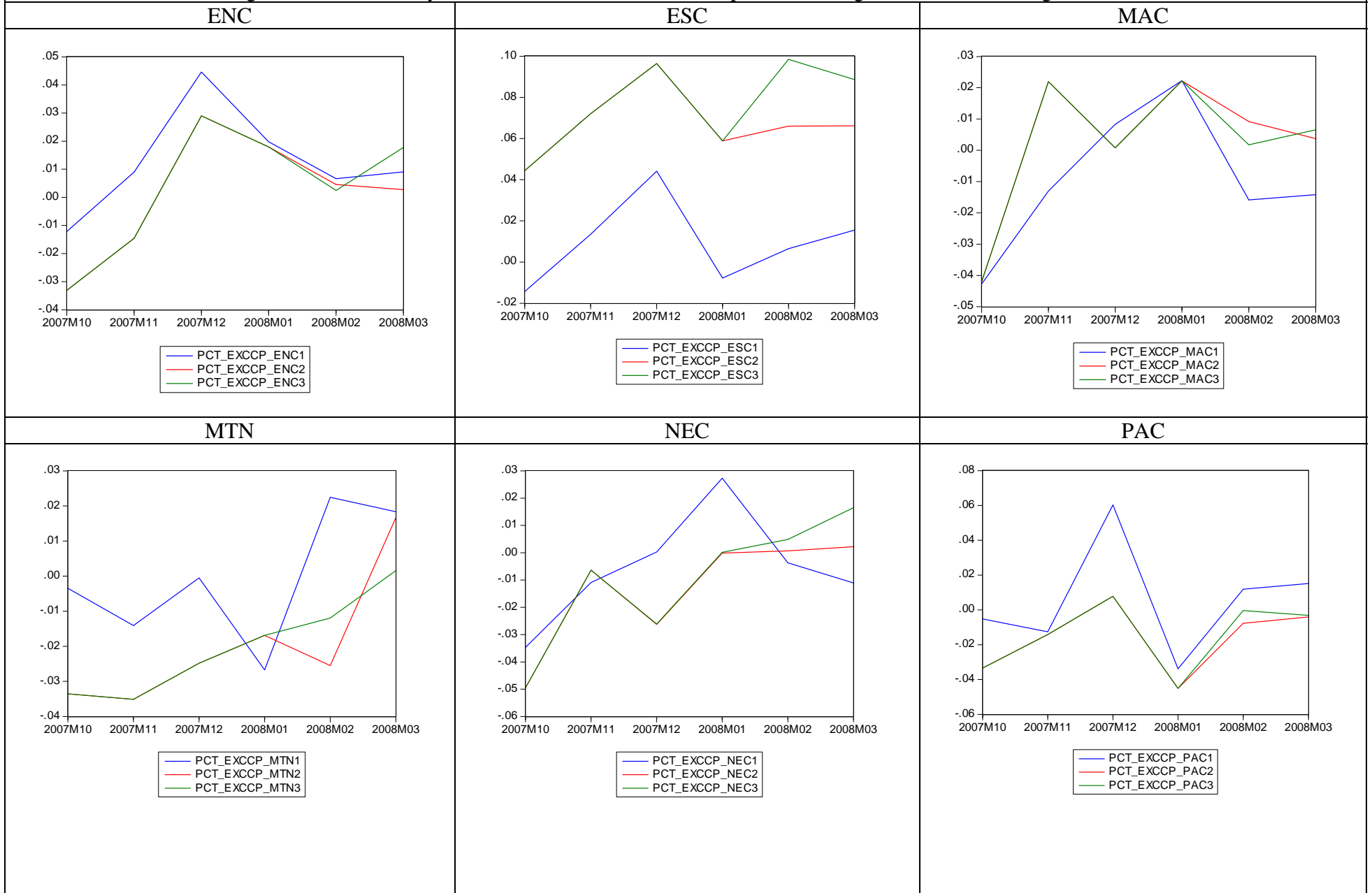
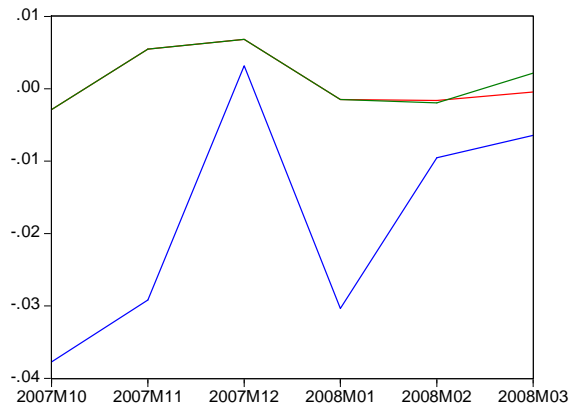


Figure II.24 (cont.)

SAC



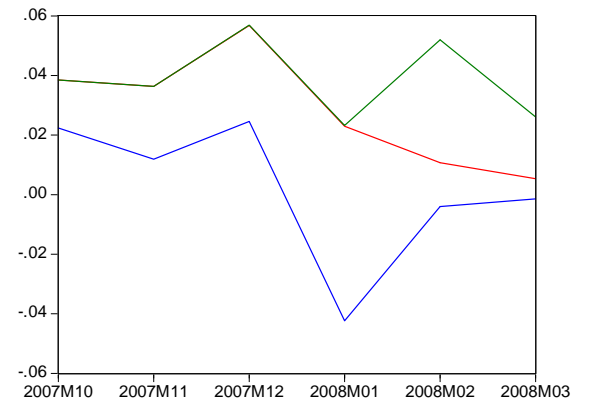
— PCT\_EXCCP\_SAC1  
 — PCT\_EXCCP\_SAC2  
 — PCT\_EXCCP\_SAC3

WNC



— PCT\_EXCCP\_WNC1  
 — PCT\_EXCCP\_WNC2  
 — PCT\_EXCCP\_WNC3

WSC



— PCT\_EXCCP\_WSC1  
 — PCT\_EXCCP\_WSC2  
 — PCT\_EXCCP\_WSC3

Figure II.25 Distillate: Residential: Out of Sample

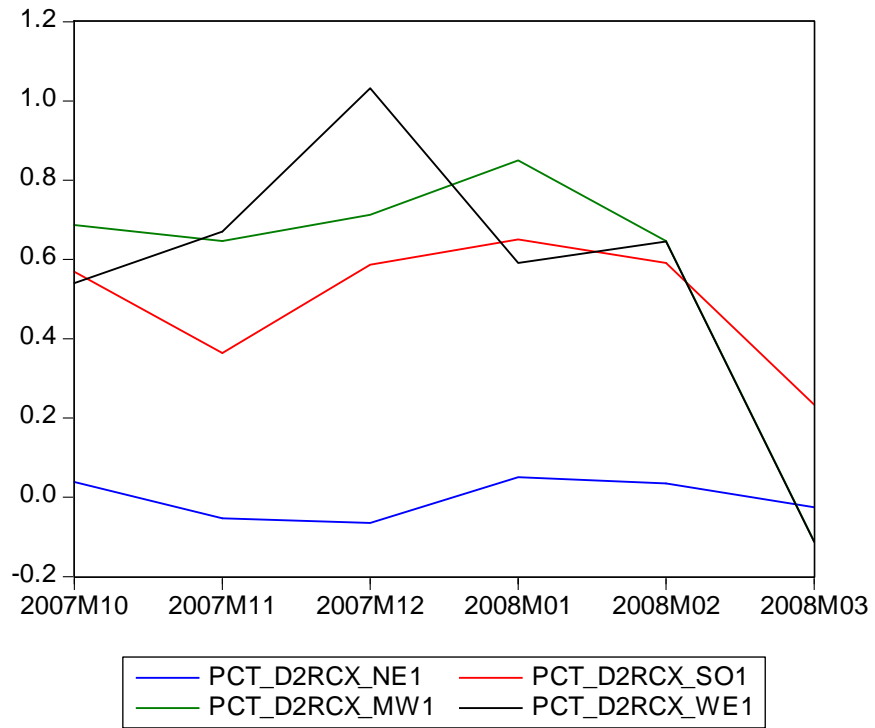


Figure II.26 Distillate: Residential: Exogenous Variables Known with Add Factors

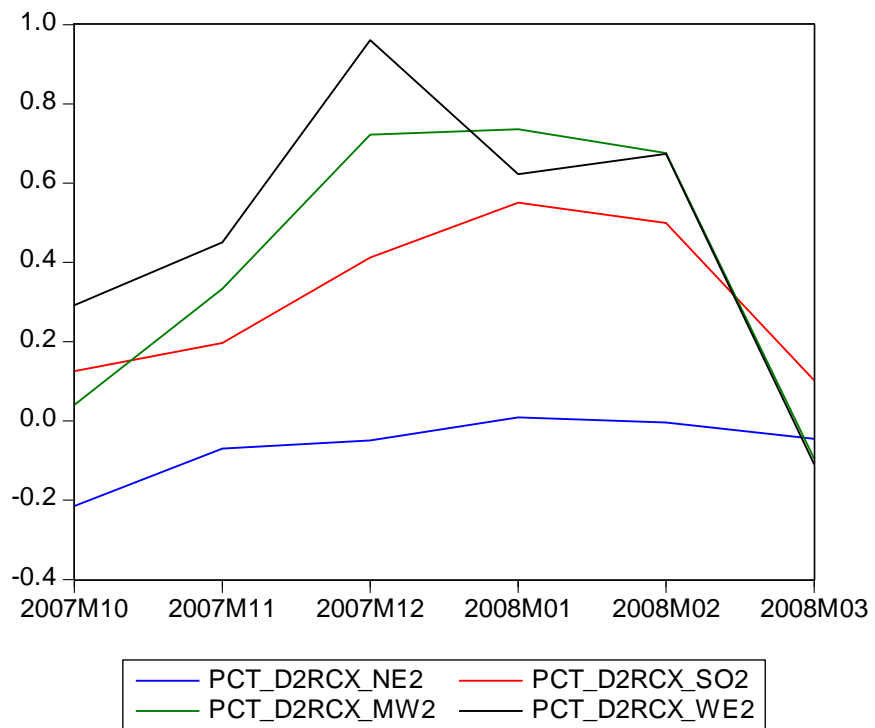


Figure II.27 Distillate: Residential: Exogenous Variables Known with No Add Factors

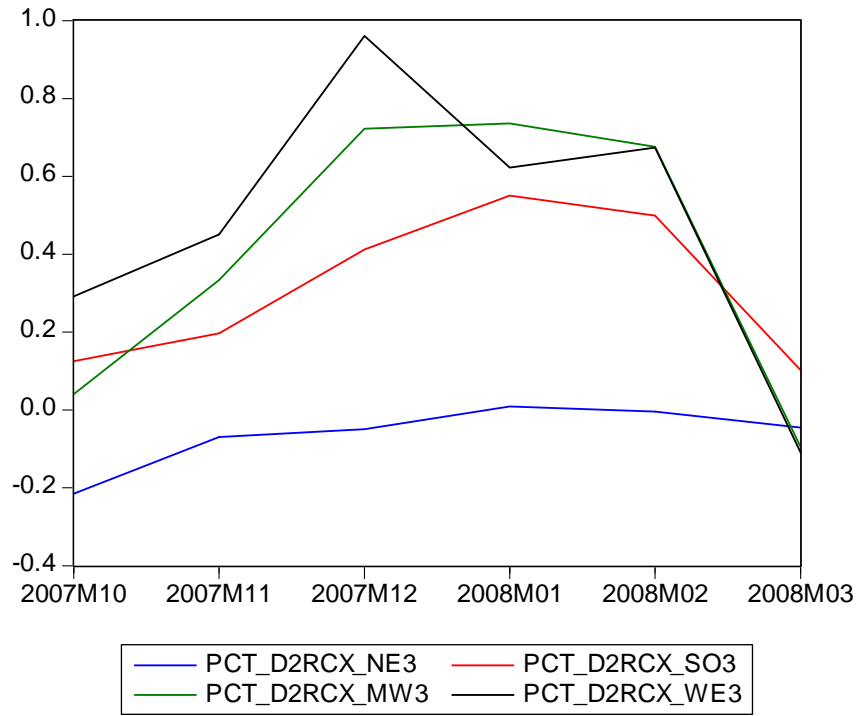
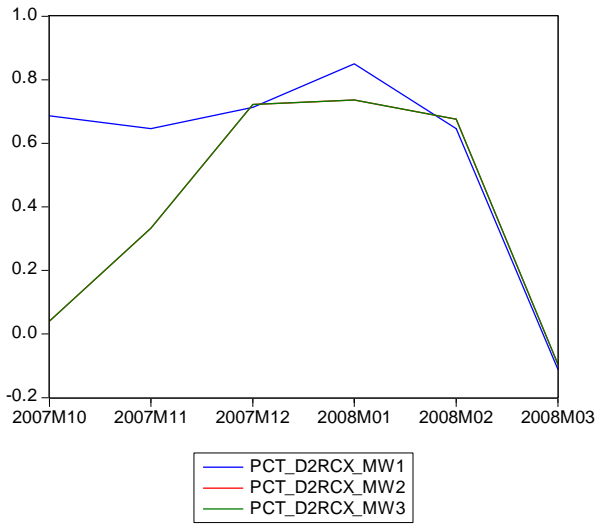
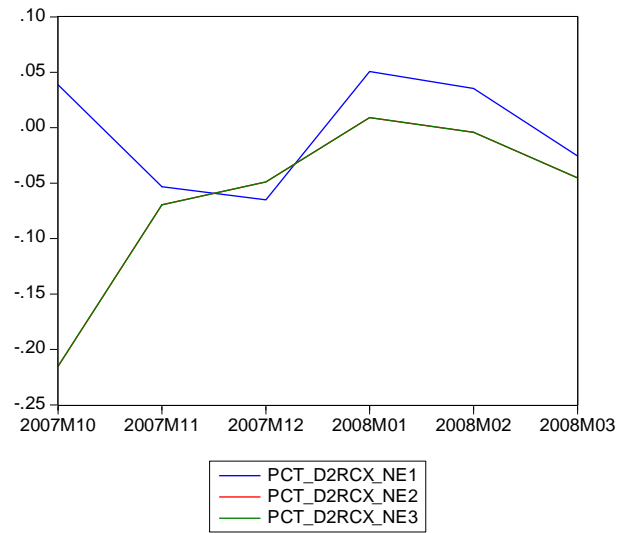


Figure II.28 Distillate: Residential: Blue=Out of Sample, Red=Exog w/ AF, Green=Exog w/out AF

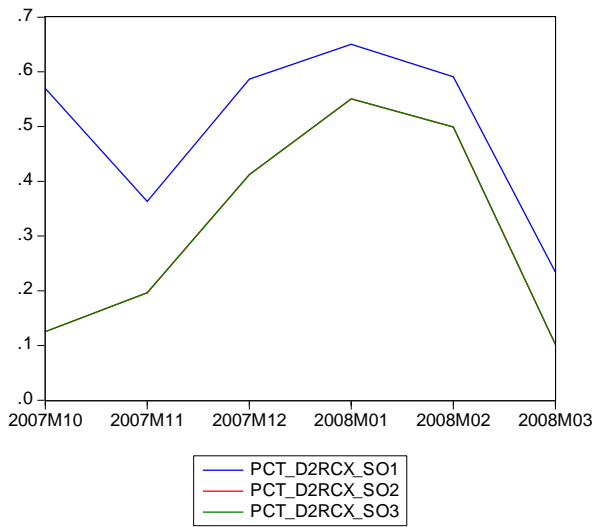
MW



NE



SO



WE

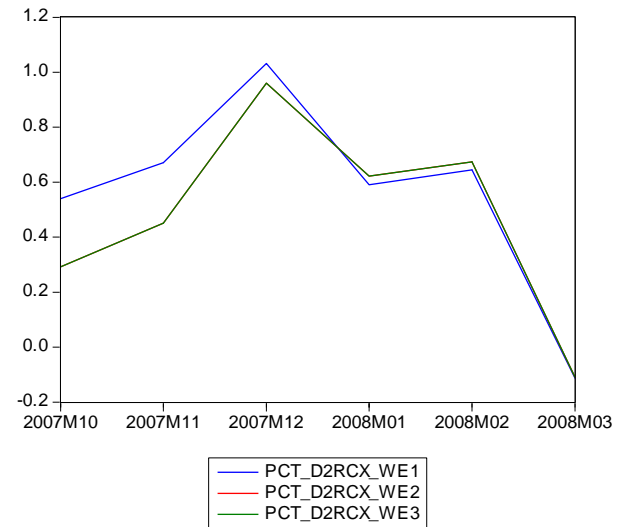


Figure II.29 Distillate: Industrial: Out of Sample

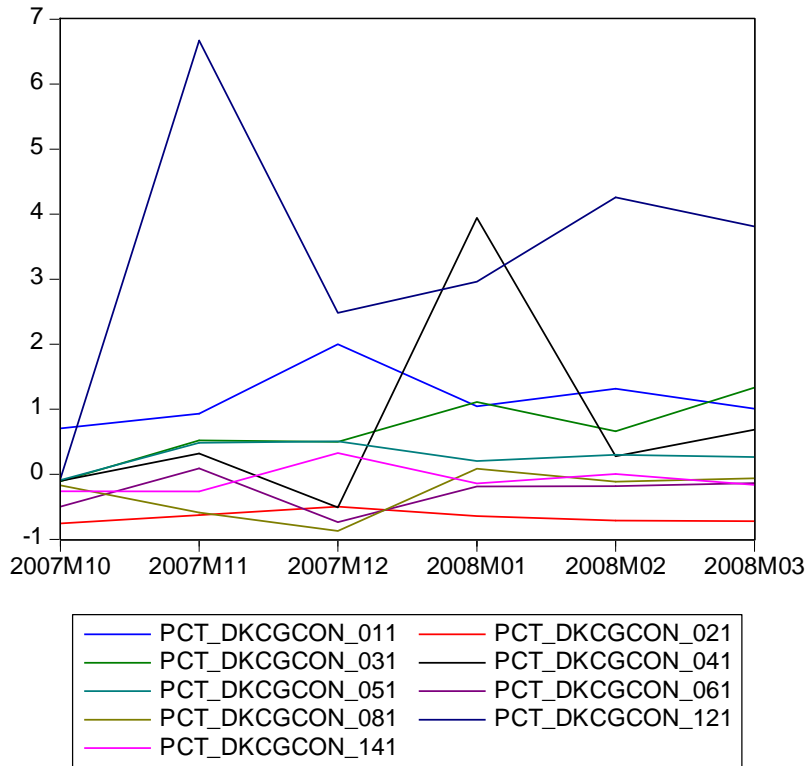


Figure II.30 Distillate: Industrial: Exogenous Variables Known with Add Factors

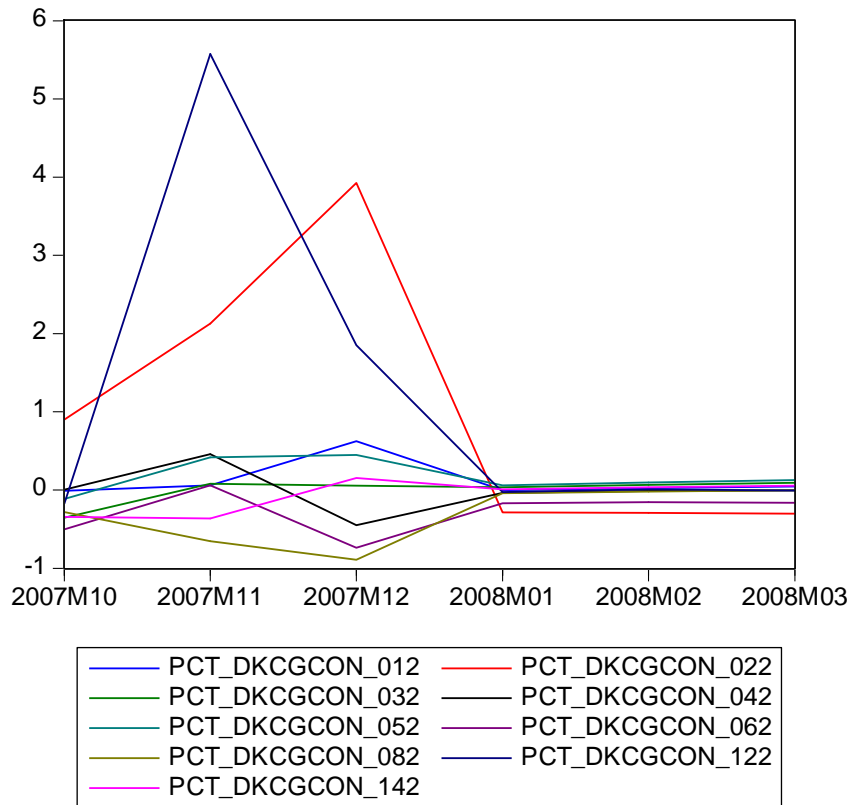


Figure II.31 Distillate: Industrial: Exogenous Variables Known with No Add Factors

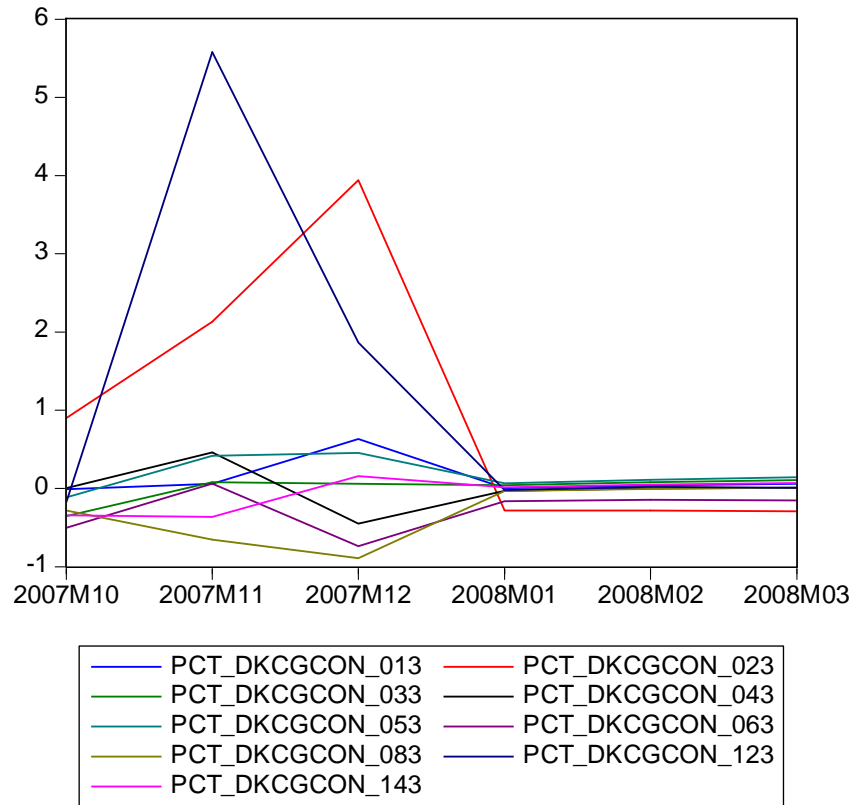


Figure II.32 Distillate: Industrial: Industrial: Blue=Out of Sample, Red=Exog w/ AF, Green=Exog w/out AF

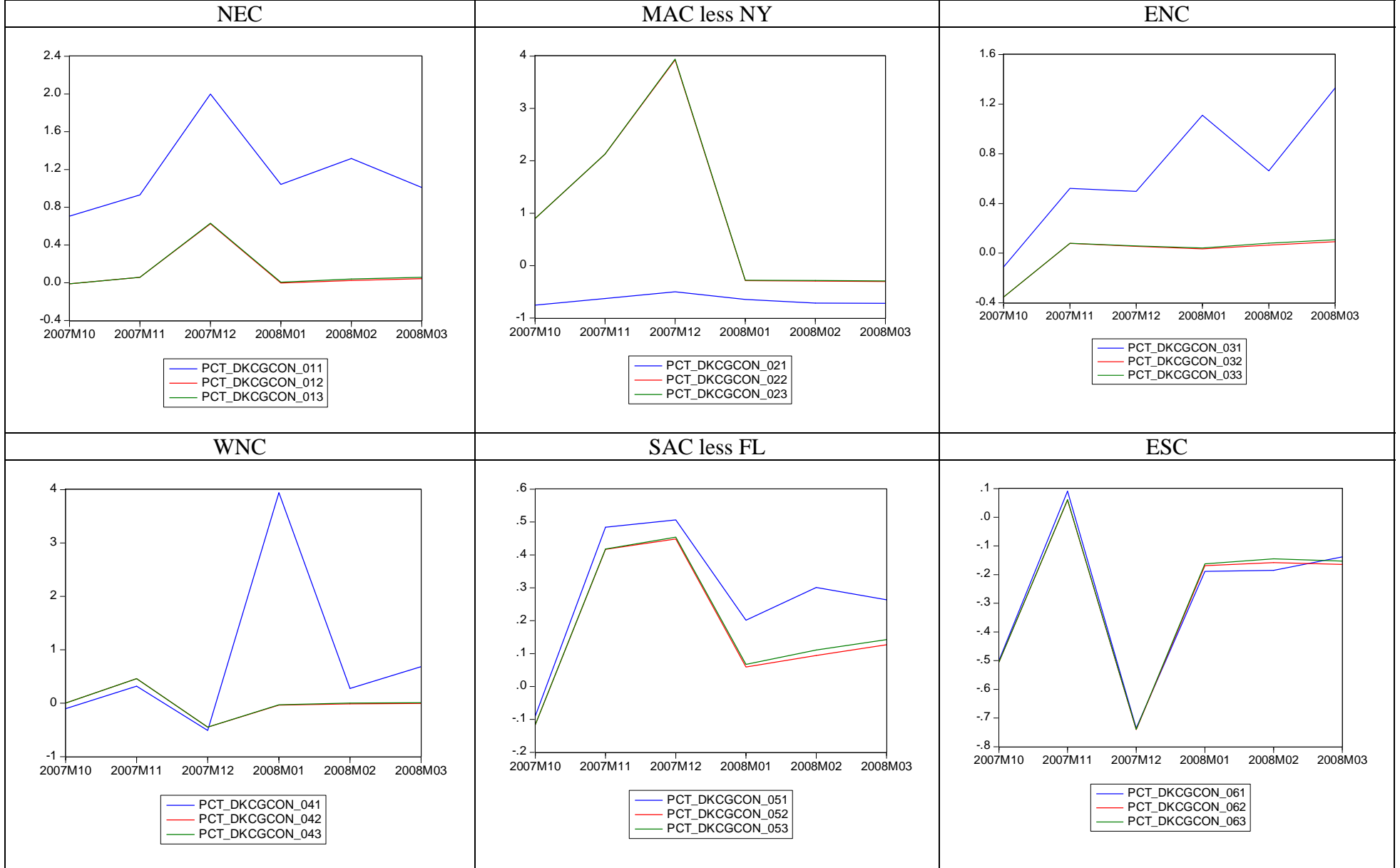
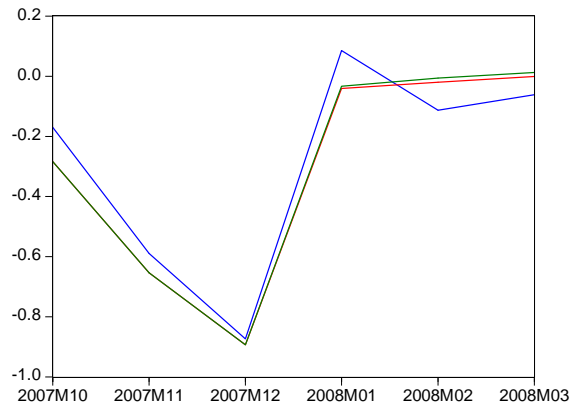


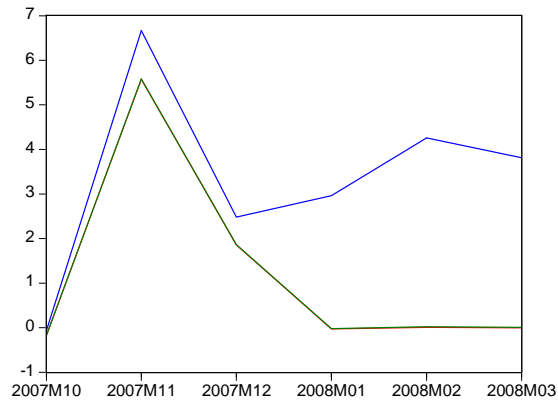
Figure II.32 (cont.)

MTN



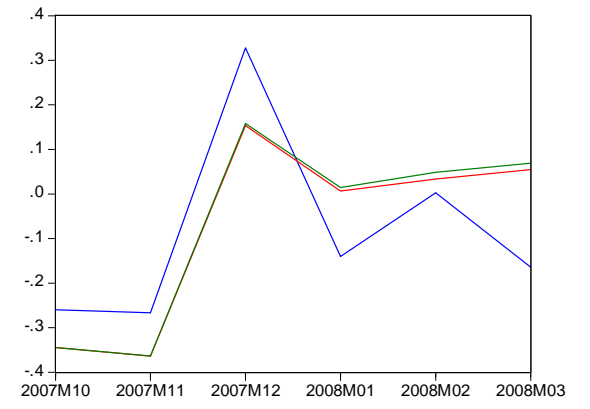
— PCT\_DKCGCON\_081  
 — PCT\_DKCGCON\_082  
 — PCT\_DKCGCON\_083

NY



— PCT\_DKCGCON\_121  
 — PCT\_DKCGCON\_122  
 — PCT\_DKCGCON\_123

Alaska & Hawaii



— PCT\_DKCGCON\_141  
 — PCT\_DKCGCON\_142  
 — PCT\_DKCGCON\_143

Figure II.33 Distillate: Commercial: Out of Sample

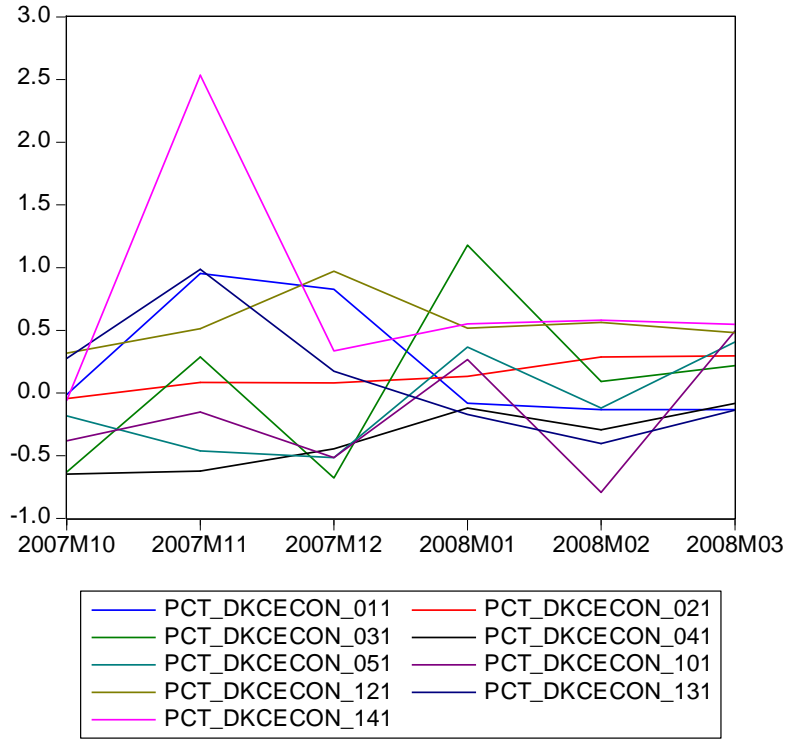


Figure II.34 Distillate: Commercial: Exogenous Variables Known with Add Factors

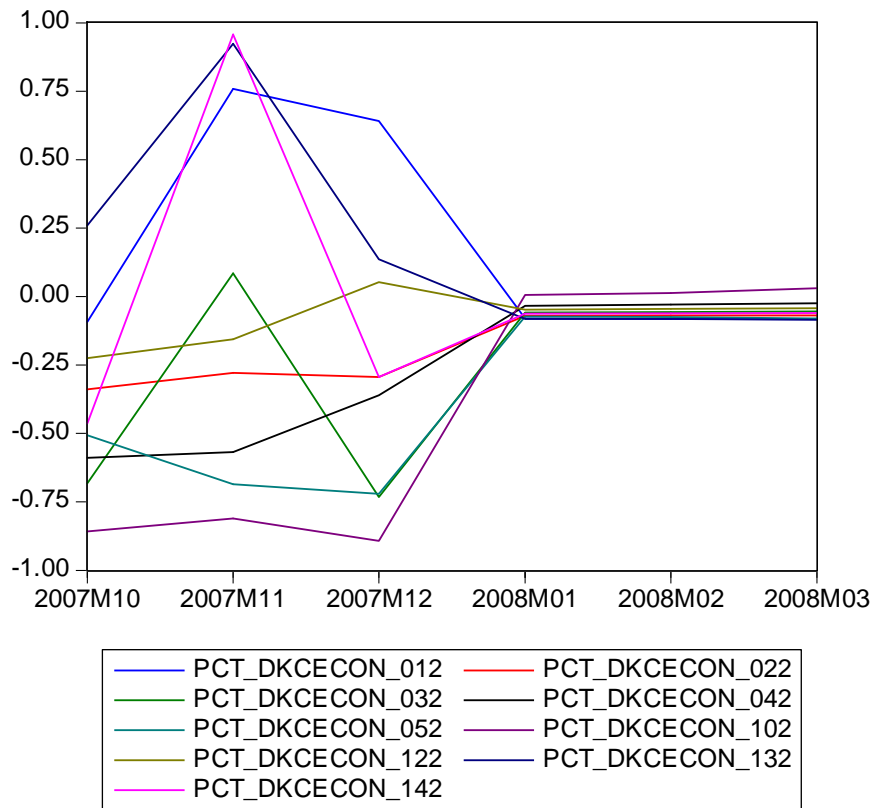


Figure II.35 Distillate: Commercial: Exogenous Variables Known with No Add Factors

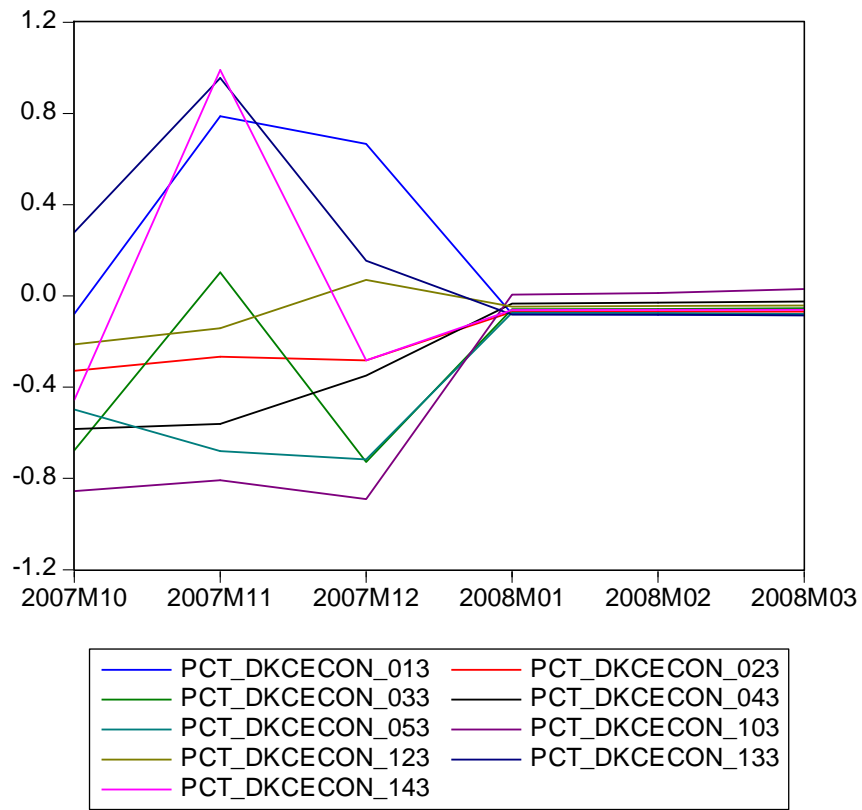


Figure II.36 Distillate: Commercial: Blue=Out of Sample, Red=Exog w/ AF, Green=Exog w/out AF

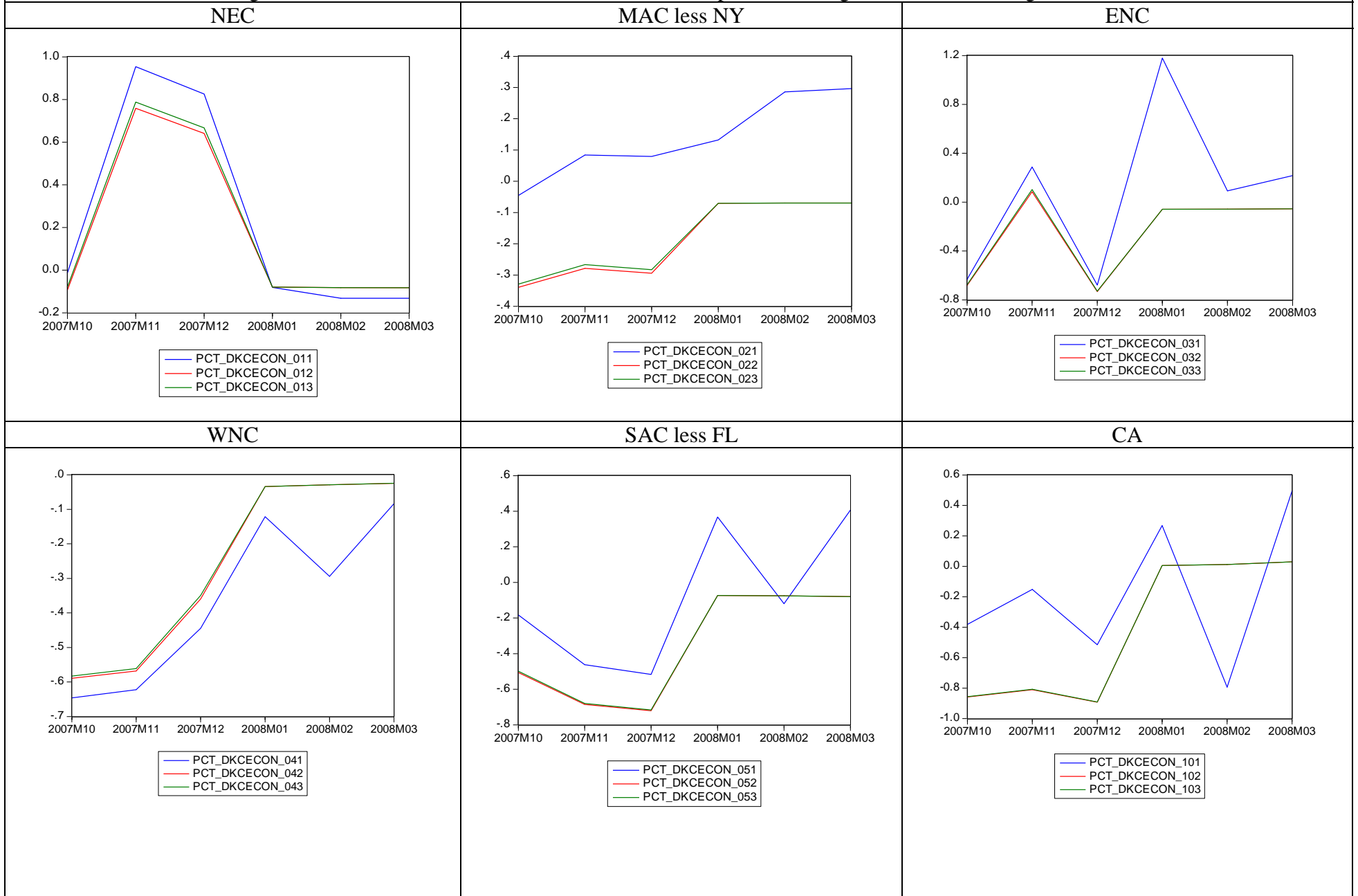


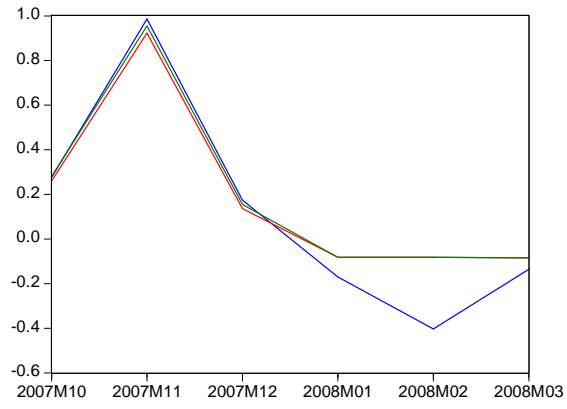
Figure II.36 (cont.)

NY



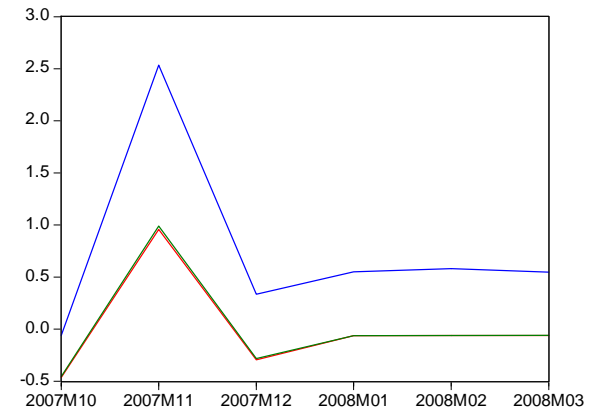
— PCT\_DKCECON\_121  
 — PCT\_DKCECON\_122  
 — PCT\_DKCECON\_123

TX



— PCT\_DKCECON\_131  
 — PCT\_DKCECON\_132  
 — PCT\_DKCECON\_133

Alaska & Hawaii



— PCT\_DKCECON\_141  
 — PCT\_DKCECON\_142  
 — PCT\_DKCECON\_143