A tree-form constant market share analysis for modelling growth causes in international trade

Feng, Yuanhua
University of Paderborn
Warburger Str. 100
33098 Paderborn, Germany
E-mail: yuanhua.feng@wiwi.upb.de

Guo, Zhichao
China Agricultural University
100083 Beijing, China
E-mail: sxguozc@hotmail.com

Peitz, Christian
University of Paderborn
33098 Paderborn, Germany
E-mail: Christian.peitz@upb.de

Tan, Xiangyong
Beijing Technology and Business University
100048 Beijing, China
E-mail: xytan@cau.edu.cn

Constant market share (CMS) analysis, also known as shift-share analysis, is an accounting method for decomposing a country’s export change into the “demand growth”, “competitive” and “interaction” effects, which was first applied to empirical studies in industrial and regional economics. CMS analysis became popular in empirical international economics with the pioneering work of Tyszynski (1951) and it has been increasingly used and refined despite continued criticism both on its theoretical and empirical aspects (Richardson, 1971, Jepma, 1986 and Milana, 1988). In recent years, in addition to its application in international economics (Chen et al., 2000 and Guo et al., 2010), CMS analysis is also widely applied in other fields, such as tourism, energy economics and firm performance. Moreover, results of the CMS model can be used for further statistical analysis (Batista, 2008) to obtain detailed information for decision making. This paper introduces a tree-form CMS for modelling trade between a focus country (A) and a destination (B) based on data classified at several levels, which extends the application spectrum of the CMS analysis clearly.

1. The tree-form CMS model

For simplicity, data classified at three levels will only be considered. That is A’s aggregate exports are first divided into n first-level categories, the i-th first-level category consists of n_j second-level sub-categories and the (i, j)-th second-level category is again composed of n_q third-level sub-categories. B’s total imports from the world are classified in the same way. Denote the quantities of A’s exports at those classification levels in the initial and final years by q^i, q^0; q^i, q^0; q^i, q^0; q^i, q^0, i = 1, …, n, j = 1, …, n_j, and k = 1, …, n_q, the quantities of B’s imports from the world by Q^i, Q^0; Q^i, Q^0; Q^i, Q^0 and Q^i, Q^0, and A’s corresponding market shares by s^0, s^i; s^0, s^i; s^0, s^i and s^0, s^i, respectively. We have q^i = ∑_i q^i = ∑_j q^j = ∑_k q^k, where ∑_i and ∑_j denote double and triple sums over i and j, and over i, j and k.

(1) Let Δ denote the change between the initial and final years. The change of A’s aggregate exports can be decomposed based on data at different levels of classification:

\[Δq = s^iΔQ^i + ΔsQ^0 + ΔsΔQ \quad \text{(overall decomposition)}\]

\[= ∑_{i,j} s^iΔQ^i + ∑_{i,j} ΔsQ^0 + ∑_{i,j} ΔsΔQ^0 \quad \text{(first-level decomposition)}\]

\[= ∑_{i,j,k} s^iΔQ^i + ∑_{i,j,k} ΔsQ^0 + ∑_{i,j,k} ΔsΔQ^0 \quad \text{(second-level decomposition)}\]

\[= ∑_{i,j,k} s^0ΔQ^0 + ∑_{i,j,k} ΔsQ^0 + ∑_{i,j,k} ΔsΔQ^0 \quad \text{(third-level decomposition)}\]

where the three terms in each row on the right-hand-side (rhs) are called the “demand growth component (Dl)”, “competitive component (Cl)” and “interaction component (Il)”, l = 0, 1, 2, 3, respectively, calculated at the overall to the third-level using the aggregated, the first-, second- and third-level classified data. These three terms at each level can be called basic decompositions of the total change. In this paper we propose to
decompose each of the three CMS components further in a symmetrical way. Based on the third-level decomposition in the fourth row of Model (1), the entire (three-level) tree-form CMS model is defined by

$$\Delta q = s^0\Delta Q + \Delta s\Delta Q + \Delta s\Delta Q$$

$$+ \left( \sum_{i,j} s^0_i \Delta Q^i - s^0_i \Delta Q \right) + \left( \sum_{i,j} \Delta s_j Q^j - \Delta s_j Q^j \right) + \left( \sum_{i,j} \Delta s_i Q^i - \Delta s_i Q^i \right)$$

$$+ \left( \sum_{i,j,k} s^0_{ijk} \Delta Q_{ijk} - s^0_{ijk} \Delta Q \right) + \left( \sum_{i,j,k} \Delta s_{ijk} Q_{ijk}^k - \Delta s_{ijk} Q_{ijk}^k \right) + \left( \sum_{i,j,k} \Delta s_{ijk} Q_{ijk}^i - \Delta s_{ijk} Q_{ijk}^i \right)$$

$$+ \left( \sum_{i,j,k} \Delta s_{ijk} Q_{ijk}^j - \Delta s_{ijk} Q_{ijk}^j \right),$$

where the three terms in the first row on the rhs of Model (2) are the overall components ($D_0$, $C_0$ and $I_0$), and those in the second, third and fourth rows are the first-, second- and third-level effects, which are called the demand growth effect ($E_1$), competitive effect ($E_2$) and interaction effect ($E_3$), respectively. Note that in this paper, the concept “component” stands for a main term according to the CMS decomposition and the concept “effect” for the difference between same-named components at two levels over each other. Also note that the sum of the level-effects at each level or within any (sub-) category is always zero, which will be confirmed by data examples in Section 3. Due to the data structure under consideration, the order of the indices $i$, $j$, $k$ and $l$ for each sum is fixed and the indices are not exchangeable. This fact rolls out a possible problem in a CMS formulation caused by the order of decomposition, because this is determined in the current context by the nature of the data. Model (2) includes all level-effects due to the classification and provides us detailed sources that cause the change of exports.

Furthermore, the $i$-th first-level branch-model is defined by taking the $i$-th element out of Model (2):

$$\Delta q_i = s^0_i \Delta Q_i + \Delta s_i \Delta Q_i$$

$$+ \left( \sum_{j} s^0_j \Delta Q^j - s^0_j \Delta Q \right) + \left( \sum_{j} \Delta s_j Q^j - \Delta s_j Q^j \right) + \left( \sum_{j} \Delta s_j Q^j - \Delta s_j Q^j \right)$$

$$+ \left( \sum_{k} s^0_{ijk} \Delta Q_{ijk} - s^0_{ijk} \Delta Q \right) + \left( \sum_{k} \Delta s_{ijk} Q_{ijk}^k - \Delta s_{ijk} Q_{ijk}^k \right) + \left( \sum_{k} \Delta s_{ijk} Q_{ijk}^i - \Delta s_{ijk} Q_{ijk}^i \right)$$

$$+ \left( \sum_{k} \Delta s_{ijk} Q_{ijk}^j - \Delta s_{ijk} Q_{ijk}^j \right),$$

where the terms in the first row on the rhs are the first-level components ($D_1$, $C_1$ and $I_1$), and those in the second and third rows are the second- and third-level effects, denoted by $E_1^i$, $E_2^i$, $E_3^i$, and $E_1^0$, $E_2^0$, $E_3^0$, respectively. We can define the $(i, j)$-th second-level branch-model by taking the $(i, j)$-th element out of (2):

$$\Delta q_{ij} = s^0_{ij} \Delta Q_{ij} + \Delta s_{ij} \Delta Q_{ij}$$

$$+ \left( \sum_{k} s^0_{ijk} \Delta Q_{ijk} - s^0_{ijk} \Delta Q \right) + \left( \sum_{k} \Delta s_{ijk} Q_{ijk}^k - \Delta s_{ijk} Q_{ijk}^k \right) + \left( \sum_{k} \Delta s_{ijk} Q_{ijk}^i - \Delta s_{ijk} Q_{ijk}^i \right)$$

$$+ \left( \sum_{k} \Delta s_{ijk} Q_{ijk}^j - \Delta s_{ijk} Q_{ijk}^j \right),$$

where the three terms in the first row on the rhs are the second-level components ($D_{2ij}$, $C_{2ij}$ and $I_{2ij}$), and those in the second row are the second-level effects, denoted by $E_1^i$, $E_2^i$, $E_3^i$, respectively. Finally, the $(i, j, k)$-th leaf-model can be defined for each final (leaf-) category by taking the $(i, j, k)$-th element out of Model (2):

$$\Delta q_{ijk} = s^0_{ijk} \Delta Q_{ijk} + \Delta s_{ijk} \Delta Q_{ijk}$$

$$+ \left( \sum_{k} s^0_{ijk} \Delta Q_{ijk} - s^0_{ijk} \Delta Q \right) + \left( \sum_{k} \Delta s_{ijk} Q_{ijk}^k - \Delta s_{ijk} Q_{ijk}^k \right) + \left( \sum_{k} \Delta s_{ijk} Q_{ijk}^i - \Delta s_{ijk} Q_{ijk}^i \right)$$

$$+ \left( \sum_{k} \Delta s_{ijk} Q_{ijk}^j - \Delta s_{ijk} Q_{ijk}^j \right),$$

where the terms on the rhs are the third-level components ($D_{3ijk}$, $C_{3ijk}$ and $I_{3ijk}$). Note that all of the branch- and leaf-models are parts of Model (2). Particularly, each branch-model looks like another tree-form CMS model defined based on data from that level to the final categories. In summary, Models (2) to (5) constitute a collection of the CMS models at different levels, which consists of a large amount of information and has a wide application spectrum. For further extensions of the tree-form CMS model see Feng et al. (2011).

2. Properties of the tree-form CMS model

Now we will discuss the sources of the level effects. The following theorem provides conditions under which
some or all of the level-effects $E_1^c$, $E_2^c$ and $E_3^c$ vanish. Proofs of the results may be found in Feng et al. (2011).

**Theorem 1:** Under the regularity conditions $q_i^c, q_i^s, Q_i^c$ and $Q_i^s > 0$, we have

Case 1: a) If $s_i^c = s_i^s = ... = s_n^c$, then $E_i^c$ vanishes;

b) If $s_i^c = s_i^s = ... = s_n^c$, then $E_i^c$ vanishes and
c) If both conditions in a) and b) are fulfilled, then $E_i^c, E_i^s$ and $E_i^c$ all vanish.

Case 2: If $Q_i^c : Q_i^s = Q_i^c : Q_i^s = ... = Q_i^c : Q_i^s$, then all of the three first-level effects vanish.

Theorem 1 reveals a more deep relationship between different variables so that the level-effects vanish, which can also be extended to effects at other levels. All conditions in Cases 1 and 2 are sufficient, but may be unnecessary. In the following we will call $q_i^c$ and $q_i^s$ endogenous and $Q_i^c$ and $Q_i^s$ exogenous quantities, which reflect A’s export structure and its changes, and B’s market structure and its changes, respectively. Conditions in Case 1 show that if the market shares are the same for all commodities in the initial and final years, respectively, all of the three level-effects caused by the classification will vanish. These conditions can be regarded as suitable mixed conditions, because the market shares depend on $q_i^c$ and $q_i^s$, and $Q_i^c$ and $Q_i^s$, simultaneously, which indicate that the inhomogeneity of A’s market share in different commodities is one of the sources of the level-effects. The homogeneity assumption on A’s market share is irrelevant in practice. Conditions in Case 2 are suitable exogenous conditions, which only depend on B’s market situation and demonstrate that, if the market growth in each of the $n$ categories is (about) the same, all of the three level-effects will be (about) zero. Now the change of exports is mainly reflected by the overall components. The inhomogeneity of the market growth in different commodities is another source of the level-effects. Data examples in Section 3 show that the market growth is sometimes roughly homogeneous.

The special case with $n = 2$ is taken to show some details. Note that now the sufficient condition in Theorem 2 of Theorem 1 becomes $Q_1^c Q_1^s = Q_2^c Q_2^s$. Let iff stand for if and only if.

**Corollary 1:** Assume that the regularity conditions of Theorem 1 hold. For $n = 2$ we have

a) $E_1^D > 0 (=0 or >0)$, iff $Q_1^c Q_2^c - Q_1^s Q_2^s (s_1^c - s_1^s) < 0 (=0 or >0)$;

b) $E_2^C > 0 (=0 or <0)$, iff $Q_1^c Q_2^c - Q_1^s Q_2^s (s_1^c - s_1^s) > 0 (=0 or <0)$;

c) $E_i^c > 0 (=0 or <0)$, iff $Q_1^c Q_2^c - Q_1^s Q_2^s (s_i^c - s_i^s) Q_i^c - (s_i^c - s_i^s) Q^0 > 0 (=0 or <0)$ and

d) The effects $E_1^c, E_2^c$ and $E_i^c$ all vanish, iff $Q_1^c Q_2^c = Q_1^c Q_2^c$ or $s_i^c = s_i^c$ and $s_i^s = s_i^s$.

Corollary 1d) shows that, for $n = 2$, conditions of Theorem 1 such that the three terms all vanish are also necessary. Corollaries 1 a) to c) indicate further when the level-effects $E_i^c$, $E_i^s$ and $E_i^c$ will be positive, zero or negative. Once one of the three terms vanishes, the other two are equal in absolute value and opposite in sign. Furthermore, if $(s_i^c - s_i^s) Q_i^c - (s_i^c - s_i^s) Q_i^s = 0$ but $Q_i^c Q_i^s - Q_i^c Q_i^s \neq 0$, $s_i^c \neq s_i^s$ and $s_i^s \neq s_i^s$, then the first-level interaction effect $E_i^c$ is fully divided into $E_i^c$ and $E_i^c$ with weights determined by $Q_i^j$ and $Q_i^j$. For $n > 2$, necessary conditions such that $E_i^c, E_i^s$ and $E_i^c$ all vanish become more complex and will not be discussed.

3. Decomposing growth causes of China’s exports to Germany

Data downloaded from the United Nations Commodity Trade Statistics Database (UN Comtrade) are used as examples. In this section, results for the periods from 2000 to 2001, 2001 to 2002, 2007 to 2008 and 2008 to 2009 will be discussed in detail. Those periods are chosen around China’s accession to WTO and the 2008 financial crisis, which have strongly affected the trade relationship between China and Germany (Guo et al., 2010). According to the Standard International Trade Classification (SITC Rev. 3, shortly SITC), the total exports are firstly divided into two first-level categories, i.e. agricultural and industrial products. Based on the 1-digit SITC, agricultural products are composed of four categories (SITC 0, 1, 2 and 4), and industrial products also consist of four categories (SITC 5-8), which are the eight second-level categories considered. SITC 3 and 9 are excluded due to many missing values. Then based on the 2-digit SITC, the four agricultural categories are composed of ten, two, nine and three, and the four industrial categories are divided into nine, nine, nine and eight sub-categories, respectively. These are regarded as the third-level categories.
At first, the results of the three-level tree-form CMS model, Model (2), are listed in Table 1, which show that each sum of the effects at the first-, second- and third-level is zero. For instance, in China’s exports to Germany from 2008 to 2009, the three third-level effects are 1214, -1612 and 398 million US dollars, respectively, which sum up to zero. Results of Corollary 1 can be now confirmed numerically. Firstly, the three first-level effects of China’s exports to Germany from 2000 to 2001 are all about zero. According to Corollary 1 c), this can only happen if \( Q_1^s : Q_2^s : Q_3^s \) or \( s_1^s = s_2^s \) and \( s_1^s = s_3^s \). After detailed checking, we find \( s_1^s \neq s_2^s \) and \( s_1^s \neq s_3^s \), but \( Q_1^s : Q_2^s = 509.482 \approx 1.06 : 1 \) and \( Q_1^s : Q_2^s = 370.613487 \approx 1.06 : 1 \) and conditions of Corollary 1 d) are roughly fulfilled. By China’s exports to Germany from 2007 to 2008 the first-level demand growth effect \( (E_1^D) \) is negative and the first-level competitive effect \( (E_1^C) \) is positive. The reasons for this are \( Q_1^s : Q_2^s : Q_3^s \) \( s_1^s - s_2^s \) \( s_1^s - s_3^s \) > 0 and \( Q_1^s : Q_2^s : Q_3^s \) \( s_1^s - s_2^s \) > 0, according to Corollaries 1 a) and b). Results in Table 1 also show that the absolute values of the second- or third-level effects may also be larger than that of the corresponding first-level effect.

### Table 1: Results of the entire three-level tree-form CMS model ($US million$)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Effect</td>
<td>D</td>
<td>C</td>
<td>I</td>
<td>D</td>
</tr>
<tr>
<td>Overall</td>
<td>573</td>
<td>-95.0</td>
<td>-5.9</td>
<td>195</td>
</tr>
<tr>
<td>( E_1 )</td>
<td>4.3</td>
<td>-3.9</td>
<td>-0.46</td>
<td>-22.7</td>
</tr>
<tr>
<td>( E_2 )</td>
<td>-61.1</td>
<td>57.4</td>
<td>3.7</td>
<td>-19.1</td>
</tr>
<tr>
<td>( E_3 )</td>
<td>85.3</td>
<td>-77.2</td>
<td>-8.1</td>
<td>-71.2</td>
</tr>
</tbody>
</table>

Because the total exports consist of two first-level categories, i.e. agricultural and industrial products, we have two first-level branch-models. We will focus on agricultural products. Corresponding results following Model (3) are given in Table 2. Table 2 confirms again that \( E_1^D + E_2^D + E_3^D = 0 \) and \( E_1^C + E_2^C + E_3^C = 0 \) hold. It is found that all of the three second-level effects of China’s exports to Germany in agricultural products from 2001 to 2002 are relatively very small, because Germany’s market growth in the four sub-categories of agricultural products in that period was roughly homogeneous with \( Q_1^s : Q_2^s : Q_3^s : Q_4^s \approx 1:0.17:0.55:0.04 \) and \( Q_1^s : Q_2^s : Q_3^s : Q_4^s \approx 1:0.17:0.52:0.04 \). That is, the condition of Case 2 in Theorem 1 is approximately fulfilled.

### Table 2: Results of the first-level branch-model for agricultural products ($US million$)

<table>
<thead>
<tr>
<th>Effect</th>
<th>( D_1 )</th>
<th>( C_1 )</th>
<th>( I_1 )</th>
<th>( E_2^D )</th>
<th>( E_2^C )</th>
<th>( E_3^D )</th>
<th>( E_3^C )</th>
<th>( E_4^D )</th>
<th>( E_4^C )</th>
</tr>
</thead>
<tbody>
<tr>
<td>00-01</td>
<td>26.2</td>
<td>21.0</td>
<td>1.2</td>
<td>-0.17</td>
<td>-3.0</td>
<td>3.2</td>
<td>12.9</td>
<td>-11.9</td>
<td>-0.97</td>
</tr>
<tr>
<td>01-02</td>
<td>25.9</td>
<td>-27.4</td>
<td>-1.4</td>
<td>-1.2</td>
<td>1.2</td>
<td>0.04</td>
<td>-4.5</td>
<td>4.7</td>
<td>-0.2</td>
</tr>
<tr>
<td>07-08</td>
<td>201</td>
<td>72.3</td>
<td>10.0</td>
<td>9.3</td>
<td>-6.7</td>
<td>-2.7</td>
<td>-33.5</td>
<td>33.1</td>
<td>0.45</td>
</tr>
<tr>
<td>08-09</td>
<td>-303</td>
<td>174</td>
<td>-30.3</td>
<td>26.4</td>
<td>-14.6</td>
<td>-11.7</td>
<td>135</td>
<td>-176</td>
<td>40.6</td>
</tr>
</tbody>
</table>

Both agricultural and industrial products are composed of four second-level categories, so there are totally eight second-level branch-models. According to the 1-digit SITC, we have chosen food and live animals (0), and machinery and transport equipment (7) as examples. Results of these two chosen second-level branch-models are shown in Table 3. It is clear that after China’s accession to WTO, the competitive component had a great drop from 36.2 to -20.4 million US dollars for food and live animals, which however exhibited a huge increment from 298 to 1145 million US dollars for machinery and transport equipment. This fact indicates that for export competitiveness, China’s accession to WTO had a negative short-term impact on agricultural products but a positive impact on industrial products.
Finally, two 2-digit SITC commodities, i.e. fish, etc. (03) and office machines, etc. (75) are chosen to show the results of the leaf-models. Results for these two models are provided in Table 4. The total export change of fish and office machines are -6.32 and -2920 million US dollars from 2008 to 2009, respectively, caused by the financial crisis. The large reduction is mostly caused by the demand growth component, because the import demand in Germany decreased significantly. The 2008 financial crisis had a clearly negative short-term impact on China’s exports of agricultural and industrial products. We see, each leaf-model based on each third-level category can still provide us very detailed information that causes the change of exports.

Table 4: Results of the chosen leaf-models ($US million)

<table>
<thead>
<tr>
<th>Name</th>
<th>Fish, crustaceans, mollusks, etc.</th>
<th>Office and automatic data-processing machines</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effect</td>
<td>$D_{3113}$</td>
<td>$C_{3113}$</td>
</tr>
<tr>
<td>00-01</td>
<td>15.9</td>
<td>10.1</td>
</tr>
<tr>
<td>01-02</td>
<td>-2.5</td>
<td>-8.3</td>
</tr>
<tr>
<td>07-08</td>
<td>23.9</td>
<td>84.1</td>
</tr>
<tr>
<td>08-09</td>
<td>-3.6</td>
<td>-2.7</td>
</tr>
</tbody>
</table>

4. Further analysis of the CMS outputs

Now we will use the decompositions for industrial products to show that it is worthy to analyze the CMS results further. The demand growth, competitive and interaction components in this class from 1994 to 2009 are displayed in Figure 1. We see both the demand growth and interaction components exhibited a great drop, while the competitive component and demand growth component, where $t = 1, \ldots, 15$, $D^*_t = D_t = 0$ for $t<10$, $D^*_t = 1$ and $D_t = t$ for $t \geq 10$. We see the structural break in the demand growth and competitive components is a rate-shift, while the interaction component exhibits a level-shift. Note that the last model is equivalent to the result selected by maximizing the $F$-value, if an ANOVA is applied to all pairs of sub-series divided by a rolling time point.
Figure 1: Developments of the demand growth, competitive and interaction components ($US million).

We can also see that the development of the demand growth component between 1994 and 2008 is relatively regular, the variation in the interaction component between 2003 and 2008 is clearly larger than before and that in the competitive component is very large. Moreover, competitive components seem to correlate to each other in a negative way. This means that if China’s competitiveness increased strongly in one year, its increment in the next year tends to be smaller. Results of the Durbin-Watson test and the correlogram show that this negative correlation is however insignificant at the 5%-level.

5. Final remarks
This paper introduces a tree-form CMS model and provides a deep discussion on its application. It is clear that the tree-form CMS model can be easily extended to multi-level classified data (Feng et al., 2011). It is also indicated that, when using the CMS model, not only the final but also the intermediate results contain useful information for decision making. Sources of effects caused by each level of classification are discussed in detail. It is found that if the market shares for all commodities are both homogenous in the initial and final years, respectively, or if the market growth in all of the categories is the same, the three level-effects will all vanish. The tree-form CMS model is applied to analyze the growth causes of China’s exports to Germany, particularly in agricultural products. These data examples also confirmed our theoretical findings. Furthermore, it is found that the growth causes before and after China’s accession to WTO (or the 2008 financial crisis) are clearly different. Finally, analysis of the outputs of the branch-model for industrial products shows that China’s accession to WTO has had a positive long-term impact on all of the three CMS components. In summary, despite a huge number of theoretical studies on the CMS model and its wide applicability, there still seems to be a big play room for the further development of theory and practice of the constant market share analysis. The current paper may open a new research direction in this area.

REFERENCES (RÉFÉRENCES)