Applying the Bass model to pharmaceuticals in emerging markets

Daniel Porath
Fachhochschule Mainz – University of Applied Science
Christian Schaefer
Boehringer Ingelheim GmbH

Albeit the Bass model was not designed for predicting sales of newly launched drugs, pharmaceutical companies commonly use it for this purpose, mainly because of its good predictive power. Empirical experience, however, mainly refers to mature markets and it is unclear how the model behaves in emerging markets. We try to fill this gap in the literature by comparing the estimation results of the Bass model between emerging markets and mature markets in a big dataset including more than 5,000 new launches from different countries. Our results show a good performance of the model in emerging markets. Compared to mature markets the estimated parameters on average are the same, but there is a higher heterogeneity between individual countries. Our findings favour the application of the Bass model in emerging markets, but also highlight the importance of selecting individual parameters for each country and therapeutic class.

Introduction

The Bass model is one of the most frequently used models for predicting the diffusion of new products. Moreover, it is one of the most studied and applied models in marketing at all. Although in the last decades there has been an uncountable number of further developments and refinements, the model in its original form (Bass 1969) still constitutes the fundamental point of reference for many authors and in many applications within different industries.

Pharmaceutical companies also intensively use the Bass model for predicting the uptake of pipeline projects and newly developed products.
Their particular interest derives from the need to predict the path of the sales volume from the time when the product is launched until saturation. Assumptions about the shape of the uptake curve need to be taken, for example, because the potential speed of market diffusion is a critical success factor for economic project evaluations by net present value, for calculations or even early operational capacity planning.

The advantage of the Bass model in this context lies in the combination of a parsimonious structure and a good (historical) fit to real data. The parsimony of the model makes it extremely easy to obtain predictions, because only three parameters have to be known: the innovation coefficient ($p$), the imitation coefficient ($q$) and the market potential ($m$). With these parameters the models takes the form

\[
n(t) = pm + (q - p)N(t) - \frac{q}{m}N(t)^2
\]  

where $t$ stands for time, $n(t)$ for the number of adoptions at time $t$, $N(t)$ for the cumulated number of adoptions up to $t$. The parameter $m$ is directly linked to $N(t)$, because it is the long-run cumulated number of adoptions.

In many applications the number of parameters even reduces to two, because estimations of the market potential normally are available from the business plan. In this case the model can be divided by $m$ and transforms to

\[
f(t) = p + (q - p)F(t) - qF(t)^2
\]  

with $f(t) = n(t)/m$ and $F(t) = N(t)/m$. With (2) the adoption in percentage of the market potential, $f(t)$, can be predicted when knowing the values of $p$ and $q$. Here forecasters typically refer to the uptake profile of analogue products, where historical data are available or to some rule of thumb deriving from past experience.

When using the Bass model in the pharmaceutical industry there are two major departures from the theoretical model. First, the variable for adoptions, $n(t)$, in (1) is replaced by sales. The volume of adoptions and sales is not the same for pharmaceuticals, because of repeated purchase. Second, the Bass model was designed for completely new products, ignoring the possible competition between different producers. In this spirit a new product in the pharmaceutical industry would be a new therapeutic class. In reality, however, the model is also used for predicting
diffusion of single brands, which may be substitutes of competing brands in the same therapeutic class. Despite these drawbacks, leading researchers in applied pharmaceutical forecasting, like Johnson (2005, p. 125), point out that the Bass model is the most commonly used model in the pharmaceutical industry. Our own practical experience within this industry supports Johnson’s observation. The prominent role of the Bass model can be justified with the empirical finding of a good fit to historical volume sales data, see e.g. Johnson (2005, p. 125).

Relying on a good historical fit, however, becomes difficult when shifting to new markets. Emerging markets, in particular, may behave differently from mature markets for many reasons, like lower absolute average GDP per capita combined with a higher growth rate, minor public health insurance systems, high co-payment by the patient for pharmaceuticals, or lower density of doctors per capita. There is relatively little literature on the diffusion of pharmaceutical drugs in emerging markets. Desiraju et al. (2004) compare ten developed to five developing countries, with the logistic model proposed by Van den Bulte (2000). Analysing one specific therapeutic class, they find some substantial differences in speed of diffusion and penetration level. Berndt et al. (2007) compare the diffusion of products from three therapeutic classes in 15 different countries, and find different effects from promotion and prices between the countries. Vakratsas and Kolsarici (2008) develop a new diffusion model for pharmaceutical products incorporating a dual-market structure, and fit the model to a specific therapeutic class. All of these studies highlight different aspects of the diffusion process of pharmaceuticals without answering the question whether the model commonly used by the industry is adequate in emerging markets and how it should be adapted compared to mature markets.

In this paper we compare the diffusion process of pharmaceuticals in mature markets and emerging markets. Unlike other studies, we focus on the Bass model as the industrial standard of the pharmaceutical industry. In contrast to the literature cited above, our comparison is not restricted to single or few therapeutic classes. In particular, we compare the performance of the Bass model in mature markets and emerging markets broadly over nine different therapeutic classes. We analyse whether the Bass model can be applied to emerging markets at all and, if so, whether and how the parameters have to be changed. We do not a priori assume that the countries with emerging markets are homogeneous, but explicitly allow for country-specific heterogeneity. In particular, we test the following hypotheses.
**H1:** The Bass model is as adequate in emerging markets (EM) as in mature markets (MM) to predict historical uptakes.

**H2:** The Bass model’s fit of historical uptakes is equal in all countries with EM.

**H3:** The innovation parameter $p$ and the imitation parameter $q$ are equal in EM and MM.

**H4:** The innovation parameter $p$ and the imitation parameter $q$ are equal in all single countries with EM.

To test these hypotheses we start from equation (2) because, as stated above, $m$ is normally determined by a different procedure. This allows us to compare EM and MM prescinding from levels of sales, but rather focusing on the shape of the diffusion process determined by the parameters $p$ and $q$.

**Methods**

Hypotheses 1 to 4 will be analysed by estimating the Bass model in the form of (2) with historical data on sales of newly launched pharmaceuticals from different countries.

We use the non-linear least square estimation technique proposed by Srinivasan and Mason (1986), and subsequently applied in many other studies (see Van den Bulte and Lilien (1997) for a discussion). Following this approach, an estimable equation can be derived by solving the differential equation (2) and regarding its first difference (see the Appendix for technical details). At a first stage we can estimate individual innovation parameters $p$ and imitation parameters $q$ for each country and each product (equation (3) in the Appendix). Apart from that, pooling the data and introducing dummy variables provides us with estimates on different levels of aggregation (see equation (4)). These include:

- average $p$ and $q$ in mature markets, and their average differences $p_E$ and $q_E$ in emerging markets
- country averages of $p$ and $q$
- average $p$ and $q$ within each therapeutic class in mature markets and their average differences in emerging markets ($p_E$ and $q_E$).

Estimating $p$ and $q$ on different levels of aggregation is necessary for
testing our hypotheses: the individual parameters $p$ and $q$ estimated at the first stage will mainly be used for comparing the performance of the Bass model in both market types (H1 and H2), but also for gaining a first impression as to whether these parameters have different values in both types of market (H3). Estimates of the average differences between the parameters in emerging markets and mature markets ($p_E$ and $q_E$) will also be used to test H3: a statistical test on H3 translates into a test on the null hypothesis that the $p_E$ and $q_E$ are zero. H4, on the other hand, can be tested by estimating the country averages of $p$ and of $q$, and comparing them within both market types. Finally, since pooling the data of different products could lead to unobserved heterogeneity with its negative effects on the estimations of $p$ and $q$, we estimate averages for individual therapeutic classes.

**Data**

The data taken from the IMS MIDAS Quantum database (see IMS MIDAS Quantum 2011) comprise quarterly volume sales of individual newly launched pharmaceuticals in 17 different countries and nine different therapeutic ATC-classes (Anatomical Therapeutic Chemical Classification System, see WHO 2011) ranging from 1999 to 2010. The sales data are projections of a continual, periodic statistically representative IMS market survey, measuring total purchases of pharmaceuticals by wholesalers, retail pharmacies and, for most countries, by hospitals. Sales are measured in standard units (SU) of a product. These are the number of dose units, like number of tablets or number of 10 mg doses of a liquid pharmaceutical sold for a particular product. Standard units are defined for all product forms, such as liquids and solids. Using standard units makes the demand for different products, offered in different pack sizes and/or formulations, comparable. Since generally the products were launched in different periods and not in all countries, the panel of data are highly unbalanced. In order to avoid seasonality we transform the series to annual sales.

The left-hand-side variable in the estimation equations ($f_{cit}$ in equations (3) and (4) of the Appendix) represents the adoptions in percentage of market potential. This variable can be calculated by dividing SU by the market potential of each product in each country. As a proxy for the market potential we use the cumulated SU observed in the last year of the sample. Generally, the quality of this proxy will be poor for pharmaceuticals launched in later periods. For this reason we exclude time series with less than five years of observations.
The countries represent nine mature markets (MM), which are Australia, Canada, France, Germany, Italy, Japan, Spain, the UK and the US, and eight emerging markets (EM), which are Brazil, China, India, Korea, Mexico, Russia, Turkey and Venezuela. The following variety of nine ATC-classes have been investigated: Diabetes (A10), Antithrombotic agents (B1), Agents acting on the renin-angiotensin system (C9), Antivirals (J5), Cytostatic (L1), Endocrine therapy (L2), Immunosuppressive agents (L4), Anti-Parkinson’s drugs (N4) and Anti-asthmatics (R3). Table 1 reports the number of products per country and therapeutic classes. The dataset includes a total of 5,076 time series where 2,895 series are from EM and 2,181 series are from MM.

Table 1  Number of products per country and therapeutic class included in the IMS MIDAS Quantum dataset

<table>
<thead>
<tr>
<th>Country</th>
<th>A10</th>
<th>B1</th>
<th>C9</th>
<th>J5</th>
<th>L1</th>
<th>L2</th>
<th>L4</th>
<th>N4</th>
<th>R3</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>20</td>
<td>3</td>
<td>26</td>
<td>12</td>
<td>57</td>
<td>6</td>
<td>9</td>
<td>6</td>
<td>10</td>
<td>149</td>
</tr>
<tr>
<td>Brazil</td>
<td>59</td>
<td>6</td>
<td>125</td>
<td>22</td>
<td>38</td>
<td>18</td>
<td>6</td>
<td>11</td>
<td>48</td>
<td>333</td>
</tr>
<tr>
<td>Canada</td>
<td>24</td>
<td>10</td>
<td>21</td>
<td>11</td>
<td>22</td>
<td>9</td>
<td>11</td>
<td>5</td>
<td>7</td>
<td>120</td>
</tr>
<tr>
<td>China</td>
<td>81</td>
<td>16</td>
<td>48</td>
<td>92</td>
<td>116</td>
<td>18</td>
<td>15</td>
<td>4</td>
<td>42</td>
<td>432</td>
</tr>
<tr>
<td>France</td>
<td>44</td>
<td>2</td>
<td>75</td>
<td>19</td>
<td>40</td>
<td>25</td>
<td>9</td>
<td>8</td>
<td>31</td>
<td>253</td>
</tr>
<tr>
<td>Germany</td>
<td>38</td>
<td>4</td>
<td>132</td>
<td>11</td>
<td>76</td>
<td>9</td>
<td>18</td>
<td>22</td>
<td>24</td>
<td>334</td>
</tr>
<tr>
<td>India</td>
<td>418</td>
<td>24</td>
<td>272</td>
<td>19</td>
<td>57</td>
<td>26</td>
<td>26</td>
<td>23</td>
<td>94</td>
<td>959</td>
</tr>
<tr>
<td>Italy</td>
<td>31</td>
<td>16</td>
<td>31</td>
<td>31</td>
<td>45</td>
<td>18</td>
<td>9</td>
<td>11</td>
<td>52</td>
<td>244</td>
</tr>
<tr>
<td>Japan</td>
<td>61</td>
<td>27</td>
<td>50</td>
<td>24</td>
<td>36</td>
<td>8</td>
<td>9</td>
<td>13</td>
<td>46</td>
<td>274</td>
</tr>
<tr>
<td>Korea</td>
<td>153</td>
<td>12</td>
<td>95</td>
<td>58</td>
<td>31</td>
<td>14</td>
<td>10</td>
<td>24</td>
<td>83</td>
<td>480</td>
</tr>
<tr>
<td>Mexico</td>
<td>61</td>
<td>2</td>
<td>41</td>
<td>18</td>
<td>43</td>
<td>10</td>
<td>13</td>
<td>7</td>
<td>20</td>
<td>215</td>
</tr>
<tr>
<td>Russia</td>
<td>26</td>
<td>8</td>
<td>54</td>
<td>29</td>
<td>64</td>
<td>19</td>
<td>6</td>
<td>4</td>
<td>15</td>
<td>225</td>
</tr>
<tr>
<td>Spain</td>
<td>25</td>
<td>4</td>
<td>88</td>
<td>22</td>
<td>41</td>
<td>15</td>
<td>10</td>
<td>10</td>
<td>18</td>
<td>233</td>
</tr>
<tr>
<td>Turkey</td>
<td>23</td>
<td>4</td>
<td>24</td>
<td>11</td>
<td>26</td>
<td>7</td>
<td>12</td>
<td>7</td>
<td>14</td>
<td>128</td>
</tr>
<tr>
<td>UK</td>
<td>34</td>
<td>12</td>
<td>44</td>
<td>13</td>
<td>68</td>
<td>13</td>
<td>8</td>
<td>16</td>
<td>22</td>
<td>230</td>
</tr>
<tr>
<td>US</td>
<td>63</td>
<td>15</td>
<td>81</td>
<td>24</td>
<td>77</td>
<td>19</td>
<td>18</td>
<td>19</td>
<td>28</td>
<td>344</td>
</tr>
<tr>
<td>Venezuela</td>
<td>31</td>
<td>2</td>
<td>41</td>
<td>4</td>
<td>1</td>
<td>8</td>
<td>3</td>
<td>3</td>
<td>30</td>
<td>123</td>
</tr>
<tr>
<td>Total</td>
<td>1,192</td>
<td>167</td>
<td>1,248</td>
<td>420</td>
<td>838</td>
<td>242</td>
<td>192</td>
<td>193</td>
<td>584</td>
<td>5,076</td>
</tr>
</tbody>
</table>

Estimation results

The adequacy of the Bass model compared in both types of market

For comparing the adequacy of the Bass model in emerging markets (EM) to mature markets (MM) according to H1, we estimate the model for
all countries and products separately (equation (3)) and then compare the results between both market types. We argue that the Bass model is adequate if both of the two conditions are met: (1) the parameters \( p \) and \( q \) fulfil the restrictions \( 0 < p < 1 \) and \( 0 < q < 1 \) imposed by the model and (2) the estimated model has a high fit to historical data in terms of the coefficient of determination, \( R^2 \).

However, if \( p \) and/or \( q \) are out of the range imposed by the model, this does not necessarily mean that the Bass model is inadequate. For some countries and products the time series are extremely short. It has been pointed out that, in short series, the parameter estimations of the Bass model tend to be biased – see, for example, Van den Bulte and Lilien (1997). The bias occurs because in short time series the market potential \( m \) used for the calculation of the left-hand-side variable (\( f_{cit} \) in equation (4)) often is misspecified. As we estimate \( m \) with the cumulated SU of the last period available, there will be a bias whenever the market potential is not yet reached by the end of period 2010. This kind of bias tends to decrease with an increasing length of the time series, and for this reason we excluded time series shorter than five years. Obviously, this threshold is arbitrary, and it will only reduce, but not avoid, the bias, and there may be even differences between MM and EM. Keeping in mind that the bias is present in all series where the market potential is not (approximately) reached by the end of 2010 and that products in EM usually have a launch delay of between one and four years compared to the first global launch date in an MM (see IMS LEAP Study 2010), we expect a higher relative amount of younger products within the EM samples, compared to the MM’s ones. Therefore, not controlling for the lengths of the time series can be misleading when comparing MM and EM.

Table 2 reports the number of estimations with at least one violated parameter restriction, and its proportion to the total number of estimations of the same lengths and type of market. Interestingly, for both types of market, the rate of estimations with violated restrictions starts from a high level for short time series and then almost monotonically decreases with an increasing length of the time series. This finding presents strong evidence for the presence of a bias in small series. When comparing EM and MM, we find higher rates of inadequate models in EM for all lengths of time series. However, the overall difference of around 10% points clearly decreases in longer series, therefore parts of it also have to be attributed to the small-series bias. The presence of this bias makes it difficult to judge in how many cases the Bass model is inappropriate. Arguing that the bias
is smallest in the longer time series (say 10, 11 or 12 years) leads to the conclusion that the model is inappropriate in not more than 9% of all launches in MM and in not more than ca. 15% of the launches in EM. This finding indicates that the Bass model is adequate for predicting sales in most series of both types of market with a slightly lower performance in EM.

In the following we will exclude all series where the parameter restrictions are violated, because we assume that most of these violations are caused by the small-sample bias. For the remaining series, we compare the (adjusted) $R^2$ of the estimations in both types of market (see Table 3).

Interestingly, $R^2$ is high in both types of market: within the EM, $R^2$ exceeds 91% in 50% of all estimations and, within the MM, the median is 92%. All percentiles are slightly higher in MM, indicating a better fit in these markets, but the difference is small. Overall the findings of this section support the hypothesis that the Bass model constitutes an adequate framework for modelling the uptake of pharmaceuticals and that the performance in EM is only slightly worse than that in MM.
Comparison of estimated parameters between both types of market

The estimations of individual countries and products (equation (3)) can also be used to compare the values of the parameter estimations between EM and MM. Again, we include only estimations meeting the restrictions imposed on the parameters by the Bass model. As shown in Table 4, the median of $p$ is virtually the same in both types of market, with 2.34% in EM and 2.37% in MM. The same result holds for $q$. Here the median is 63.21% in EM and 62.18% in MM. The difference in $p$ increases for the higher percentiles indicates that the distribution of $p$ is positively skewed in EM. For both coefficients $p$ and $q$, the fluctuation seems to be higher in EM, because the 10%-percentile falls below and the 90%-percentile exceeds that of MM.

The comparison of the point estimators in Table 4 is only indicative, because the sampling error of each estimated coefficient is ignored. In order to compare the average innovation and imitation parameter in both types of market, we pool the data and estimate (4). The main results are provided in Table 5. As before, $p$ and $q$ are the estimated innovation and imitation parameter for MM, whereas $p_E$ and $q_E$ stand for the differences to these parameters in EM: we find that the average innovation parameter is 3.89% in MM and 3.97% in EM. The high $p$-value of $p_E$ indicates that the

<table>
<thead>
<tr>
<th>Percentiles</th>
<th>Innovation parameter ($p$)</th>
<th>Imitation parameter ($q$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EM</td>
<td>MM</td>
</tr>
<tr>
<td>10</td>
<td>0.00238</td>
<td>0.00337</td>
</tr>
<tr>
<td>25</td>
<td>0.00848</td>
<td>0.00981</td>
</tr>
<tr>
<td>50</td>
<td>0.02339</td>
<td>0.02366</td>
</tr>
<tr>
<td>75</td>
<td>0.05093</td>
<td>0.04858</td>
</tr>
<tr>
<td>90</td>
<td>0.09957</td>
<td>0.08251</td>
</tr>
<tr>
<td>No. of series</td>
<td>1,688</td>
<td>1,489</td>
</tr>
</tbody>
</table>

Notes: Estimation of equation (3); EM = emerging markets; MM = mature markets

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>$t$ ratio</th>
<th>$p$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p$</td>
<td>0.03887</td>
<td>51.97</td>
</tr>
<tr>
<td>$p_E$</td>
<td>0.00078</td>
<td>0.76</td>
</tr>
<tr>
<td>$q$</td>
<td>0.44811</td>
<td>82.83</td>
</tr>
<tr>
<td>$q_E$</td>
<td>-0.00106</td>
<td>-0.14</td>
</tr>
</tbody>
</table>

Note: Pooled estimation of equation (4); number of observations = 27,887, $R^2 = 73.82\%$
difference is not significant for any commonly used level of significance. The average imitation coefficient is 44.81% in MM and 44.71% in EM. Again this difference is clearly insignificant. Therefore, on average there is no empirical evidence for different innovation or imitation parameters in EM.

Comparison of estimated parameters within both types of market

The estimation of the average parameters can be affected by heterogeneity among the countries within the same type of market. For example, the single EM countries could all (or partly) have different innovation and imitation coefficients compared to MM, but these differences could cancel each other out in the pooled data. If there is country-specific heterogeneity, H4 has to be rejected and the assumption of equal coefficients in all countries of the same type of market is too restrictive.

To analyse H4, we estimate country-specific parameters for EM and then statistically test the null hypothesis that these parameters are equal in all countries. The first column of Table 6 reports the $p$-values of this test.

For both parameters the null hypothesis is clearly rejected. As can be seen from the other columns of Table 6, this result does not alter when excluding single countries or when forming regional clusters: $p$-values remain low when excluding single countries (Turkey, Brazil, India) from the estimation or when regarding regional clusters (China, India, Korea or Brazil, Mexico, Venezuela).

The heterogeneity between the countries with EM raises the question whether the countries with MM are more homogeneous concerning $p$ and $q$. To answer this question, we also estimate country-specific models for MM and then test the joint hypothesis that the coefficients are equal for

<table>
<thead>
<tr>
<th>Table 6</th>
<th>$p$ values for joint hypotheses of equal coefficients in EM</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p$ values for the null hypothesis that all coefficients are equal in</td>
<td></td>
</tr>
<tr>
<td></td>
<td>All EM countries</td>
</tr>
<tr>
<td>Innovation coefficients</td>
<td>0.0000</td>
</tr>
<tr>
<td>Imitation coefficients</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Note: Pooled estimation of (4) with additional dummies for each country
all countries. As can be seen from the \( p \)-values in the first column of Table 7, the hypothesis can be rejected only for the innovation coefficient. When excluding Japan, the joint hypothesis of equal parameters in all countries (see Table 7) cannot be rejected on a 1% level. Without Japan and the US, the \( p \)-value exceeds 10% for both parameters.

To sum up, we find some significant differences between the coefficients within the subgroups of markets. In MM these differences are mostly caused by Japan and – to a lower degree – by the US. The assumption of equal parameters within MM therefore seems appropriate when excluding Japan. This exclusion can also be justified by the monitoring regulations in Japan, which are the most restrictive of all MM. Japanese laws exceed much more frequent and detailed monitoring by doctors when prescribing newly launched pharmaceuticals. Comparing with other countries, this might have major impact on the shape of the uptake curve, and could be reflected in different innovation and imitation parameters. Therefore the group of MM will no longer contain Japan in the following estimations. In EM the exclusion of one or two single countries does not create a homogenous cluster. As a consequence, the parameters of the Bass model used for MM cannot simply be adapted to EM with the same rule for all countries.

A second implication of this finding is that the test presented in Table 5 is not adequate with its assumption of constant \( p_E \) and \( q_E \) over all EM countries. This assumption can be relaxed by estimating country-specific values of \( p_E \) and \( q_E \). These are reported in Table 8. It should be noted that \( p_E (q_E) \) represents the difference between the innovation (imitation) parameter of the country listed in that particular row to the average \( p(q) \) in mature markets.

The average \( p \) in MM (without Japan) is 3.89% and the average \( q \) is 44.88%. To repeat: the coefficients \( p_E \) and \( q_E \) in Table 8 measure how much these values change for the country indicated in the same row, and the corresponding \( p \)-values indicate whether this change is significant. From inspecting the \( p \)-values we find no significant differences in China, Mexico

<table>
<thead>
<tr>
<th>Table 7</th>
<th>( p )-values for joint hypotheses of equal coefficients in MM</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>p values for the null hypothesis that all coefficients are equal in</strong></td>
<td><strong>All MM countries</strong></td>
</tr>
<tr>
<td><strong>Innovation coefficients</strong></td>
<td>0.0000</td>
</tr>
<tr>
<td><strong>Imitation coefficients</strong></td>
<td>0.1340</td>
</tr>
</tbody>
</table>

Note: Pooled estimation of (4) with additional dummies for each country
Applying the Bass model to pharmaceuticals in emerging markets

and Venezuela. For all other countries, at least one of both coefficients is significantly different from zero on a 10%-level of significance. On the same level, both coefficients are significantly different in Korea and Turkey, whereas in Brazil and India only $p$ is different and in Russia only $q$. In some cases, $p$ is significantly higher (India, Korea) and in other cases it is lower (Brazil and Turkey) than in MM. The same applies for $q$: we find a significantly positive difference for Turkey, and a negative difference for Korea and Russia. Overall, the results in Table 8 suggest that there are significant differences between EM and MM, which cancel each other out when taking averages over all EM. As stated before, the Bass model cannot simply be adapted to EM with just one rule. The different patterns encountered in Table 8 corroborate this finding, at the same time suggesting that, to find appropriate values of $p$ and $q$, it is even necessary to analyse each country separately.

### Controlling for therapeutic classes

The models estimated so far do not account for the fact that the data are about different products including a variety of nine different therapeutic areas. The estimations therefore could be affected by unobserved heterogeneity. Usually, the remedy against unobserved heterogeneity consists of adding dummy variables for each group, but as the dataset contains more than 2,000 different products, non-linear estimation with product dummies becomes unfeasible. On the other hand, most of the product-specific effects can be explained by the different therapeutic classes since the variation caused by the unmet medical need between the nine therapy areas should be much higher than the variation within the

<table>
<thead>
<tr>
<th>Country</th>
<th>$p$</th>
<th>$p$-value of $p$</th>
<th>$q$</th>
<th>$p$-value of $q$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brazil</td>
<td>-0.00924</td>
<td>0.000</td>
<td>0.00081</td>
<td>0.957</td>
</tr>
<tr>
<td>China</td>
<td>0.00215</td>
<td>0.360</td>
<td>0.01102</td>
<td>0.511</td>
</tr>
<tr>
<td>India</td>
<td>0.00505</td>
<td>0.000</td>
<td>0.01290</td>
<td>0.195</td>
</tr>
<tr>
<td>Korea</td>
<td>0.00475</td>
<td>0.009</td>
<td>-0.03071</td>
<td>0.012</td>
</tr>
<tr>
<td>Mexico</td>
<td>-0.00267</td>
<td>0.312</td>
<td>0.00858</td>
<td>0.668</td>
</tr>
<tr>
<td>Russia</td>
<td>-0.00035</td>
<td>0.203</td>
<td>-0.03601</td>
<td>0.061</td>
</tr>
<tr>
<td>Turkey</td>
<td>-0.01253</td>
<td>0.000</td>
<td>0.08408</td>
<td>0.000</td>
</tr>
<tr>
<td>Venezuela</td>
<td>-0.00438</td>
<td>0.122</td>
<td>-0.01092</td>
<td>0.777</td>
</tr>
</tbody>
</table>

Note: Rows represent the result from estimation of (4) with $D_x = 1$ for the country indicated in the row and otherwise zero; the coefficients of the basic category, MM (excluding Japan), are $p = 0.03887$, $q = 0.4488$. 

Table 8  Country-specific differences of $p$ and $q$ compared to MM
classes. Therefore, to account for product effects in our estimations, we re-estimate the model of Table 8 with additional therapeutic class-specific dummies, assuming that the product-specific effects within each therapeutic class can be neglected. Results can be found in Table 9.

As before, the coefficients $p_E$ and $q_E$ measure the average difference to MM, but here the therapeutic class is held constant for each comparison. The main effect from controlling for therapeutic classes is that, for Korea, $p_E$ turns out to be insignificant. All other coefficients that are significant in Table 8 remain significant and display only minor changes in size. Therefore, controlling for therapeutic classes does not alter the main results observed before: for many countries, the Bass model parameters used in MM should be adapted to EM and there is no unique rule for all countries telling which parameter needs to be changed and to what extent.

The differences reported in Table 9 are averaged over all therapeutic classes, i.e. differences between the therapeutic classes cannot be observed. In order to analyse these differences we estimate the Bass model separately for each therapeutic class and country (see Table 10 in the Appendix). In many cases there are only few products per therapeutic class, so Table 10 reports only the results for therapeutic classes with at least ten products launched in the respective country.

In Table 10 there is no country where none of the estimated $p_E$ and $q_E$ is significantly different from zero. Most notably, the variation between the therapeutic classes is high, not only for EM, but also for MM. The coefficients $p$ and $q$ for MM are shown in the first two rows of Table 10. The estimated innovation coefficients range between 1.75% (therapeutic class L4) and 4.89% (L2), and the estimated imitation coefficients are between 41.27% (N4) and 48.68% (L4). The fluctuations of the

<table>
<thead>
<tr>
<th>Country</th>
<th>$p_E$</th>
<th>$p$ value of $p_E$</th>
<th>$q_E$</th>
<th>$q$ value of $q_E$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brazil</td>
<td>−0.00712</td>
<td>0.000</td>
<td>−0.02432</td>
<td>0.111</td>
</tr>
<tr>
<td>China</td>
<td>0.00240</td>
<td>0.310</td>
<td>0.01327</td>
<td>0.439</td>
</tr>
<tr>
<td>India</td>
<td>0.00474</td>
<td>0.003</td>
<td>0.00496</td>
<td>0.661</td>
</tr>
<tr>
<td>Korea</td>
<td>0.00153</td>
<td>0.394</td>
<td>−0.03089</td>
<td>0.000</td>
</tr>
<tr>
<td>Mexico</td>
<td>−0.00283</td>
<td>0.233</td>
<td>0.01224</td>
<td>0.532</td>
</tr>
<tr>
<td>Russia</td>
<td>0.00064</td>
<td>0.818</td>
<td>−0.05024</td>
<td>0.009</td>
</tr>
<tr>
<td>Turkey</td>
<td>−0.01048</td>
<td>0.000</td>
<td>0.07795</td>
<td>0.000</td>
</tr>
<tr>
<td>Venezuela</td>
<td>−0.00493</td>
<td>0.094</td>
<td>−0.03097</td>
<td>0.157</td>
</tr>
</tbody>
</table>

Note: See Table 8, additional dummy variables included for each therapeutic class
Table 10  Estimated coefficients of country-specific parameters for each therapeutic class

<table>
<thead>
<tr>
<th></th>
<th>A10</th>
<th>B1</th>
<th>C9</th>
<th>J5</th>
<th>L1</th>
<th>L2</th>
<th>L4</th>
<th>N4</th>
<th>R3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p$</td>
<td>0.03999***</td>
<td>0.03878***</td>
<td>0.04224***</td>
<td>0.03545***</td>
<td>0.04213***</td>
<td>0.04891***</td>
<td>0.01754***</td>
<td>0.03947***</td>
<td>0.04096***</td>
</tr>
<tr>
<td>$q$</td>
<td>0.44468***</td>
<td>0.41591***</td>
<td>0.47799***</td>
<td>0.42762***</td>
<td>0.43653***</td>
<td>0.45074***</td>
<td>0.48679***</td>
<td>0.41272***</td>
<td>0.45054***</td>
</tr>
<tr>
<td>$p_E$ (Brazil)</td>
<td>-0.00762</td>
<td>-0.00979***</td>
<td>-0.00675</td>
<td>-0.01038</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.01373***</td>
</tr>
<tr>
<td>$q_E$ (Brazil)</td>
<td>0.02481</td>
<td>-0.04432*</td>
<td>0.00417</td>
<td>-0.05066</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.00573</td>
</tr>
<tr>
<td>$p_E$ (China)</td>
<td>-0.00639</td>
<td>-0.01154*</td>
<td>0.03250***</td>
<td>-0.00688*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.00031</td>
</tr>
<tr>
<td>$q_E$ (China)</td>
<td>0.02796</td>
<td>-0.02292</td>
<td>0.0099</td>
<td>0.01836</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.07077</td>
</tr>
<tr>
<td>$p_E$ (India)</td>
<td>0.00249</td>
<td>0.00580***</td>
<td>0.00542</td>
<td>0.02993***</td>
<td>0.01538***</td>
<td>-0.01320*</td>
<td>-0.00215</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$q_E$ (India)</td>
<td>0.01886</td>
<td>-0.02497</td>
<td>0.0563</td>
<td>0.00966</td>
<td>0.02089</td>
<td>0.11859*</td>
<td>-0.00071**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$p_E$ (Korea)</td>
<td>-0.00218</td>
<td>0.01149***</td>
<td>0.02045***</td>
<td>-0.01701***</td>
<td></td>
<td></td>
<td></td>
<td>-0.00469</td>
<td>0.00126</td>
</tr>
<tr>
<td>$q_E$ (Korea)</td>
<td>0.13612***</td>
<td>-0.08292***</td>
<td>-0.07350*</td>
<td>-0.01833</td>
<td></td>
<td></td>
<td></td>
<td>-0.0698</td>
<td>-0.06720**</td>
</tr>
<tr>
<td>$p_E$ (Mexico)</td>
<td>0.02369***</td>
<td>-0.02014***</td>
<td>-0.00498</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.00049*</td>
<td></td>
</tr>
<tr>
<td>$q_E$ (Mexico)</td>
<td>-0.03123</td>
<td>0.02144</td>
<td>0.07107</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.05246</td>
<td></td>
</tr>
<tr>
<td>$p_E$ (Russia)</td>
<td>-0.00274</td>
<td>-0.01016**</td>
<td>0.0105</td>
<td>0.01365**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.02810***</td>
</tr>
<tr>
<td>$q_E$ (Russia)</td>
<td>-0.00046**</td>
<td>-0.03794</td>
<td>-0.02577</td>
<td>-0.0359</td>
<td></td>
<td></td>
<td></td>
<td>0.07978</td>
<td></td>
</tr>
<tr>
<td>$p_E$ (Turkey)</td>
<td>-0.02126***</td>
<td>-0.01152**</td>
<td>-0.01421***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$q_E$ (Turkey)</td>
<td>0.09305</td>
<td>0.03623</td>
<td>0.05864</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$p_E$ (Venezuela)</td>
<td>0.00571</td>
<td>-0.01132***</td>
<td>-0.03941</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.0098</td>
<td></td>
</tr>
<tr>
<td>$q_E$ (Venezuela)</td>
<td>-0.04723</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0397</td>
<td></td>
</tr>
</tbody>
</table>

See Table 8; $p$ and $q$ refer to the group of mature markets excluding Japan
Separate estimations for each therapeutic class
The stars *, **, *** indicate that the coefficients are significantly different from zero at a 10%- , 5%- or 1%-level of significance
Results are not reported when the number of time series is less than 10
coefficients between the therapeutic classes seem to be high compared to
the fluctuation between the countries. This finding implies that adapting
the coefficients to the therapeutic classes is at least as important as adapting
them to individual countries with EM.

A rationale supporting the result of differentiating the shape of the
uptake curve for therapeutic classes is obviously a high degree of difference
in the unmet medical need between the therapeutic areas (see Trusheim
et al. 2010). For instance, in cancer therapy (ATC-classes L1 and L2),
the hope for a groundbreaking new therapy is much higher than in other
classes, like cardiovascular issues, where the cure and health status of
patients is already much better controlled.

Summary and conclusions

With this study we try to answer the question of whether it is advisable for
pharmaceutical companies to use the Bass model when predicting the sales
of newly launched products and, if so, how the innovation parameter and
the imitation parameter change versus mature markets. These questions
have important implications for practitioners because the Bass model is
the most commonly used model for predicting the uptake curve. The most
important conclusions of our findings are as follows.

In mature markets the Bass model performs well. As a rule of thumb
the sum of the model’s parameters $p$ and $q$ should be in the ballpark of
50%, with the innovation parameter being much lower than the imitation
parameter.

In mature markets the average innovation coefficient $p$ is about 4% and
the average imitation coefficient $q$ is about 45%. Discriminating between
individual therapeutic classes we find innovation parameters ranging
between 2% and 5%, and imitation parameters between 41% and 49%.
With the exception of Japan we do not find significant differences between
individual countries with mature markets. Therefore it is an adequate
strategy to use one and the same value of $p$ and $q$ for forecasting sales in
different countries with mature markets. Forecasts can be improved, though,
by adapting $p$ and $q$ to the characteristics of the individual therapeutic class.

When it comes to forecasting sales in emerging markets, the Bass model
still performs well: we find that the model provides a good description
of past uptake curves in emerging markets, with at least 85% of all the
estimations being in line with the model’s parameter restrictions and a
generally high fit. Overall, however, the performance is slightly lower than
in mature markets.
The problem of fixing the parameters of \( p \) and \( q \) for predictions in emerging markets, however, is more complex than in mature markets. Albeit on average the parameters display the same size as in mature markets \((p = 4\%, \ q = 45\%)\), in many individual countries \( p \) and \( q \) differ significantly from these values. A simple one-size-fits-all rule as in mature markets therefore seems to be inadequate. Additionally, as in mature markets, in emerging markets we also find significant differences between therapeutic classes. As a consequence, when using the Bass model for predictions in emerging markets, forecasters should consider both country-specific and therapeutic-class-specific values for \( p \) and \( q \).

There are some issues that our study does not address. For example, the estimations presented here do not account for differences between individual products. We argue, however, that the majority of these differences are captured by the therapeutic classes. Neither do we analyse whether therapeutic-class-specific effects and country-specific effects are independent. Independence would allow predictors to assess the change in \( p \) and \( q \) separately for the individual country and therapeutic class. Both accounting for product-specific heterogeneity and testing for independence fail because the non-linear least squares estimations do not converge with too many dummy variables. Addressing these issues with a different estimation technique would be an interesting issue for future research.

**Appendix: Estimation equations**

The estimation equations used in this paper are derived by solving the differential equation (2). Following Srinivasan and Mason (1986), we estimate the first difference of the solution of (2) with non-linear least square estimators. Introducing indices for each country \( c \) \((c = 1, \ldots, C)\) and each product \( i \) \((i = 1, \ldots, I)\), the estimation equation is:

\[
f_{cit} = \left( \frac{1 - \exp(-(p_{ci} + q_{ci})t)}{1 + \frac{q_{ci}}{p_{ci}} \exp(-(p_{ci} + q_{ci})t)} \right) - \frac{1 - \exp(-(p_{ci} + q_{ci})(t-1))}{1 + \frac{q_{ci}}{p_{ci}} \exp(-(p_{ci} + q_{ci})(t-1))} + \varepsilon_{cit}
\]

where \( C \) stands for the total number of countries and \( I \) for the total number of products.

For obtaining average coefficients in both market types and statistically
testing their difference, it is necessary to pool the data and introduce dummy variables: when the first \( M \) countries are MM (\( c = 1, \ldots, M \) with \( M < C \)) and the following \( M \) countries are EM (\( c = M + 1, \ldots, C \)), the dummy variable for the market type \( D_c \) can be defined as a variable taking the value zero if \( c \leq M \) (i.e. in MM) and one if \( c > M \) (i.e. in EM). The pooled model with this dummy is:

\[
f_{cit} = \left( \begin{array}{c}
1 - \exp\left( -\left( p + q + D_c \left( p_E + q_E \right) \right)t \right) \\
1 + \frac{q + D_c q_E}{p + D_c p_E} \exp\left( -\left( p + q + D_c \left( p_E + q_E \right) \right)t \right) \\
1 - \exp\left( -\left( p + q + D_c \left( p_E + q_E \right) \right)(t - 1) \right) \\
1 + \frac{q + D_c q_E}{p + D_c p_E} \exp\left( -\left( p + q + D_c \left( p_E + q_E \right) \right)(t - 1) \right)
\end{array} \right) + \varepsilon_{cit} \tag{4}
\]

Here \( p \) and \( q \) are the average innovation and imitation coefficients in MM, and \( p_E \) and \( q_E \) measure the (average) difference to these coefficients in EM. Testing H3 in this setting evolves to a test of significance on \( p_E \) and \( q_E \). For testing H4, equation (4) will be estimated with additional dummy variables for the individual countries, so that the innovation parameter and imitation parameter can vary for each country. Finally, we will also consider therapeutic-class-specific dummy variables in equation (4), so that we can check if the results are affected by heterogeneity between the different therapeutic classes.

**Acknowledgement**

The data presented here were the result of an independent analysis and do not represent IMS figures.

**References**


Applying the Bass model to pharmaceuticals in emerging markets


About the authors

Daniel Porath is professor of quantitative methods at Fachhochschule Mainz, University of Applied Sciences. His research and professional interests are related to the applicability of econometric methods in business. He has published in different refereed journals such as Journal of Banking and Finance, Journal of Financial Services Research and Empirical Economics.

Christian Schäfer is Director Global Forecasting at Boehringer Ingelheim GmbH. He holds a Masters Degree in Business Administration from University of Gävle, Sweden, a Masters Degree in Economics and a PhD from Johannes Gutenberg-University Mainz, Germany. His research interests are mainly in applied econometrics, uncertainty management and applied business forecasting.

Address correspondence to: Daniel Porath, Fachhochschule Mainz, University of Applied Science, Lucy-Hillebrand-Straße 2, 55128 Mainz, Germany.

Email: daniel.porath@wiwi.fh-mainz.de