Correlation tools

for the final finishing and testing area

Ensuring that tire testing machines read and correlate correctly is key task for any manufacturer when trying to ensure its tires meet the needs of its customers

by Dr Shaun M. Immel, Chief Technology Officer, Micro-Poise Measurement Systems, LLC, USA

At tire manufacturing facilities around the world, final finish tire testing systems will typically measure a few thousand tires each day to determine the final quality of many different types of tires. To accomplish this high volume of testing, most facilities have multiple tire uniformity, dynamic balance, and geometry measurement machines in use. At some point, a need will arise for these tire manufacturers to ensure that their final finish tire measurement machines are in agreement with each other.

Across the industry, this need is typically referred to as measurement machine correlation. Tire customers naturally have a desire to understand whether or not the quality of the tires they receive are properly and consistently evaluated when tested through different measurement machines. In order to ensure this, they expect quality measurements of a single tire taken across several different machines to "agree". Tire manufacturers call on final finish equipment suppliers to verify and guarantee their machine correlation. This paper will investigate and discuss the several tools used to evaluate final finish measurement machine correlation or, at most, what this author believes to be "measurement agreement".

"Tire manufacturers call on final finish equipment suppliers to verify and guarantee their machine correlation"

Conventional correlation procedures and tools used to study this phenomenon have limited use when applied to final finish test and measurement equipment. This is due to the fact that these tools and techniques are better suited for the study of system behaviors when the user is trying to find an inherent connection between two or more variables, perhaps to enable subsequent control of that system or for the prediction of future system behavior. What is truly desired in the previously mentioned tire testing scenario is to ensure that several machines used to test the same types of tires have measurements that are in agreement with each other. This measurement agreement is actually a special case of the more general measurement correlation.

There is one special consideration that is worth noting when thinking about the measurement agreement issue. If a set of different measurement systems (machines and processes) do not possess a high level of measurement repeatability individually, it becomes extremely difficult to evaluate their agreement. With poor repeatability, the distribution (range) of measurement is so wide that it consequently becomes difficult to truly assess measurement agreement. A generic example of this is shown in Figure 1. On the left, two probability distributions are shown (one black and one red) in which their averages differ by one unit and $\sigma = 1.25$ units. On the right, two similar probability distributions are shown in which their averages differ by the same amount and $\sigma = 0.25$ units. It is very difficult to judge whether the distributions on the left are different. However, it feels much easier to make a visual determination with the distributions on the right. The extensive distribution overlap on the left makes it difficult to ascertain agreement.

In order to address this poor repeatability issue, it is recommended that the measurement machines be tuned and maintained prior to any correlation testing. This ensures that they are as repeatable as possible. It is also advisable that multiple measurements be taken on each separate measurement machine/process, and measurement agreement analysis be performed with the averages of the results from each machine/process. This helps remove the effects of poor machine/process repeatability on the measurement agreement analysis.

Regardless of the metric used to assess measurement machine agreement, it becomes very difficult to answer the typically posed question of, "do the machines agree or not?" Each tool and technique has advantages and disadvantages in determining this agreement characteristic, but it is difficult to set limits on any numerical calculation in order to make a final go/no-go decision.

In addition, is measurement agreement really that important across the whole measurement range of the machine? Most would agree that machine measurement agreement is certainly most important in the neighborhood of any quality screening limit, and much less important for extremely high quality and extremely low quality tires, where it really does not matter if the measurements from different machines do not quite match exactly.

Another major issue encountered in the assessment of measurement machine/process agreement is the absence of a true 'NIST-traceable' standard for final finish tire measurement equipment. With most measurement devices, there is a hardened, stable, NIST-traceable standard to which any measurement device can be compared. (NIST is the National Institute of Standards and Technology, United States Department of Commerce). Final finish tire testing equipment does not have any such hardened and stable standard. There is no master polymeric

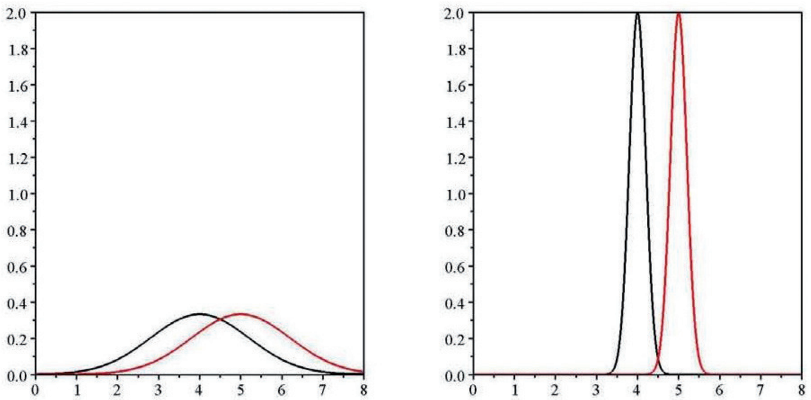

Figure 1: Measurement agreement comparison illustrating poor repeatability (left) and good repeatability (right)

tire-like device that can be used to assess measurement agreement or calibrate a final finish testing machine. As a matter of practice, 'master tires' are typically used. Master tires are tires set aside from the normal production flow, which typically have been selected for their inherent measurement stability and have been exercised for several cycles on separate equipment to further stabilize their measurements. However, the properties and measurements of these master tires change over time as they age and wear as a result of the measurement process itself.

Only by focusing on quality machine selection, proper machine set-up, alignments, maintenance, calibrations, and other similarly important factors to ensure an accurate measurement, can a true assessment of measurement agreement be performed. By far, the key to the measurement agreement equation is machine accuracy. If a tire is put into exactly the same testing conditions on two or more measurement machines, the measurements should always agree. When assessing measurement machine agreement and, more importantly, when adjusting or correcting machine agreement, the emphasis should always be placed on the accuracy of the machine as opposed to chasing master tire measurement agreement. Chasing a master tire value agreement will lead to behaviors of intentionally modifying a machine to ensure agreement with a changing target and therefore straying from the ultimate goal of truly accurate measurement.

This being said, there are many statistical tools that can be used to describe the correct process for determining machine correlation. The next sections will discuss several major tools and techniques, chosen by the author as being relevant for accurately measuring tire testing correlation

Tool A: Linear regression

The most popular tool (but not necessarily the most optimal) used to analyze machine agreement (correlation) is linear regression. This tool is based on the fundamental that when comparing measurements from two machines, or measurements taken on a machine to master tire values, a $y = x$ (least squares fit line equation) should be identified if the machine measurements agree. That is, the identified linear regression equation, in $y = m \cdot x + b$ format, should identify $m = 1$ and $b = 0$. Typically this equation is evaluated for its level of correlation by calculating the Coefficient of Determination in the linear regression model – the square of the sample correlation coefficient designated as R^2. The closer this R^2 value is to one, the higher the level of correlation is said to exist between the two sets of data.

There are several issues associated with using this particular linear regression-based metric. The correlation coefficient of a linear regression model is sensitive to several things, including, but not limited to, the spread of the data under comparison. In addition, there are very few general guidelines that indicate when the R^2 is 'good enough' to show agreement. There are different schools of thought on whether the b term should be constrained to 0 (or not) when performing the linear regression line identification. If this term is not constrained to zero, then how close to zero must it be before it is deemed acceptable? Answers will be different in either of these cases, which also make the assessment of the proximity of the slope variable m to 1 difficult to interpret. The key point here is that the spread of individual tire measurements can directly affect the outcome of the analysis, or more importantly can be used to manipulate the outcome of the study if this methodology is used exclusively. The results of this type of analysis are interesting and useful. However, the author recommends staying away from using only these results to assess the absence or achievement of measurement agreement.

Tool B: GMUTS

A well known, but not well published, automotive and tire industry tool used to assess measurement machine agreement is the General Motors Uniform Testing Scale (GMUTS). This tool, as developed by the General Motors Corporation in order to ensure consistency in testing throughout their own organization, incorporates analysis of both machine repeatability and correlation. This technique assigns three separate numerical ratings from one to 10 to a machine. The machine must rank high enough in each of these three categories in order for it to be approved for production measurement. These three areas include:

1. A linear regression equation rating – based on a fitted regression line's ability to predict the master value of each parameter under measurement most importantly near the screening limit, but also at zero and at twice the screening limit when comparing measured values or measured values with master values.
2. A separate regression rating – based on coefficient of determination (i.e. R^2).
3. A repeatability rating – based on the repeatability of measurements.

The dramatic advantages of this method over standard linear regression are that it looks at several dimensions of correlation, and built into the process are strict guidelines under which master tires may be selected. These guidelines, too numerous and detailed to cover in this paper, are specific enough to ensure that the coefficient of determination is more meaningful and that the tires are sufficiently stable to assess the machines repeatability. The fact that this method includes a repeatability assessment is extremely important. It is not worth testing machine measurement agreement if the measurement system is not considered to be repeatable and accurate. Another major advantage of this method is the emphasis on measurement agreement around the anticipated screening limits. Less emphasis is placed in measurement agreement at measurement values near zero and twice the screening limit where it is less important for making economic decisions based on product quality.

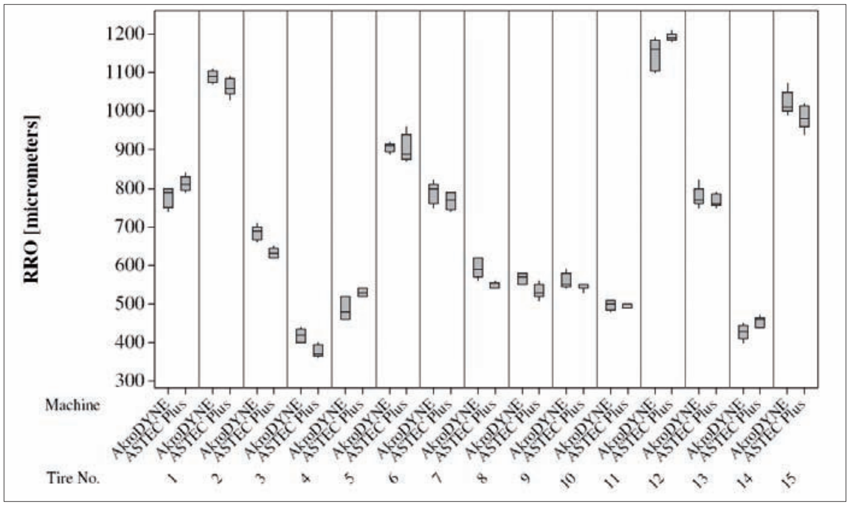

Figure 2: Boxplot of a 15 tire set of Tire Radial Runout data measured on ASTEC Plus and AkroDYNE systems

Executing the GMUTS process is quite straightforward. Several master tires or master tire and wheel assemblies are tested multiple times on the machine to be evaluated and compared with master values or on both machines if testing two machines. A modified linear regression equation is fit through the measured data. Item (1) above is a number from one to 10 assigned from a predetermined table based on the regression equation's ability to predict the screening limit. That is, after the regression equation is calculated, the equation is used to predict values at zero, at the screening limit of the particular measurement under evaluation, and at approximately twice the screening limit. If the regression equation was perfect, it would return exactly zero, the screening limit, and twice the screening limit. Based on how close these numbers are to the truth values, the rating is assigned. The additional equation used to assign the exact rating number is weighted, putting much more emphasis on the regression equation's ability to predict values around the screening limit compared with the extremes. This is philosophically in the right direction and one of the biggest advantages of this measurement agreement assessment tool. The assessment of items (2) and (3) above are straightforward. Based on the calculated R^2 value (closeness to 1.0) and a standard deviation-like calculation (closeness to 0.0), a numerical rating from one to 10 is assigned from values in a predetermined table.

Tool C: GR&R

The Gauge Reproducibility and Repeatability (GR&R) tool is typically never associated with the study of measurement agreement (correlation); however, there is a potential novel use of this tool in assessing the agreement of automated tire measurement equipment. Typically the 'reproducibility' dimension of the GR&R is used to evaluate how closely two manual inspectors acquire measurements using a particular measurement device. This same process can be used to measure how closely two automated machines are measuring (as opposed to two operators). The same basic GR&R analysis results interpretation applies. The additional benefit of this tool comes from assessing the repeatability of the measurement systems along with the agreement between machine measurements.

Tool D: Correlation coefficients

There are other correlation coefficients defined that can be used to assess measurement agreement, which are of limited use. These calculated quantities include the Spearman's or Kendall's Rank Correlation Coefficient and the Pearson's Product-Moment Correlation Coefficient. The Spearman's Rank Correlation Coefficient is a measure of statistical dependence between two variables and assesses how well that relationship can be described using a monotonic function (Anderson, 2005). The Kendall's Rank Correlation (or Kendall's Tau) Coefficient is also a measure of the association between two variables. It does this primarily by comparing the number of concordant and discordant pairs (Correlation (Pearson, Kendall, Spearman), 2009). The Pearson's correlation coefficient between two variables is defined as the covariance of the two variables divided by the product of their standard deviations and varies between -1 and 1, depending on the level of correlation similar to the other known correlation coefficients (Anderson, 2005). These tools have limited use in studying measurement agreement as they are better served in system behavior analysis when looking for general correlations between independent and dependent variables.

The 'Best' method

The several methods discussed in this paper can be used to assess measurement agreement. However, the question still remains which method is most effective? There is certainly no universally accepted answer to this question, but a strong case can be made to recommend the primary principles upon which the GMUTS tool has been created. These key points to an effective measurement agreement tool for tire testing production using GMUTS are:

1. Careful selection of stable master tires to minimize the effects that tires and tire/machine have on the measurement results (sufficient spread of individual master tire values and average values around screening limits of interest).
2. Assessment of the best fit line's ability to predict master tire values (or second machine's values) with double weighting of measurements around the screening limits of interest.
3. Assessment of the machine repeatability to ensure this does not sway the analysis results.
4. Assessment of the coefficient of determination (R^2) of the best fit line, (results are not used exclusively and selection of master tires helps to ensure meaningful interpretation of the R^2 value).

In the next section, an example measurement agreement analysis is performed using several of the tools discussed in this paper.

Example analysis

The best way to gain a deeper understanding of the numerous measurement agreement tools reviewed in this paper is to review an example analysis. Consider a case where the Radial Runout (RRO) of 15 separate tires have been measured by a TGIS-SL (Tire Geometry Inspection System with Sheet of Light) geometry system manufactured by Micro-Poise. These tires were measured by a TGIS-SL system mounted on a Micro-Poise manufactured ASTEC PLUS Tire Evaluation Center and by a separate TGIS-SL system mounted on an AkroDYNE Dynamic Balance machine. Each tire was measured a total of five times on each measurement system. Data analysis and discussion follow regarding the application

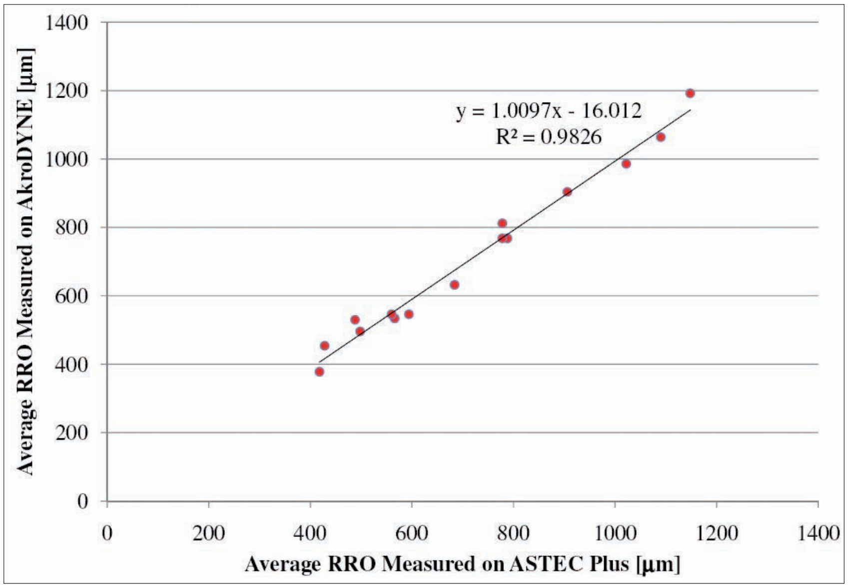

Figure 3: Linear regression plot of average tire Radial Runout data: TGIS on AkroDYNE vs TGIS on ASTEC Plus

of three previously discussed measurement agreement tools.

One of the most simple visual (non-analytical) methods to begin to assess measurement agreement between two sets of data is the boxplot. A boxplot of the 15 tire measurement data set for Radial Runout (RRO) is shown in Figure 2.

The vertical bars of the boxplot represent the distribution of the five measurements for each tire. The short horizontal lines, located near the middle of the filled rectangles, represent the averages of the five measurements. For each tire, the averages of the five measurements taken on both measurement platforms appear to be 'close in value'. As a result, without any kind of numerical evaluation, it appears from this chart that there is satisfactory agreement between the two sets of RRO data for each tire. The next logical step is to apply some of the tools described in this paper to numerically analyze measurement agreement.

Tool A application: linear regression

The linear regression plot of the RRO data shown in Figure 3 contains a best fit line with a slope of 1.0097 and a y-axis intercept of -16.012µm with a resulting coefficient of determination of R^2 = 0.9826. Generally speaking, a slope value close to 1.0000, a y-axis intercept close to 0.0000, and a coefficient of determination R^2 close to 1.0000 is desired to indicate measurement agreement. The calculated values for each of these parameters, associated with the example data set, are close to ideal but difficulty arises when determining measurement agreement. Any specification on slope, y-axis intercept, and R^2 is somewhat arbitrary with little physical significance. This process does not take into account the potential screening limit of the system. Instead it places equal emphasis across the whole spectrum of measurement values. Further analysis is required in order to help with completing this assessment.

Tool B application: GMUTS

There is no direct application of the GMUTS to RRO data as it was originally defined for uniformity data. If GMUTS RRO specifications exist, they are unknown to the author. However, some general statements can be made here based upon the principles underlying the GMUTS process. Typically, the spread of master tire values in a GMUTS system span from near zero to almost twice the typical screening limit. A typical screening limit for RRO may be approximately 1,000µm. So, by these standards, one may conclude that there is insufficient spread of the master tire values and their mean value is not close enough to the anticipated screening limit. Despite these limitations of the tire set, the R^2 value is relatively high and would rank rather highly on the GMUTS scale. The regression equation evaluated at the screening limit returns a value of y = 1.0097 × 1,000µm - 16.012µm = 993.7µm. This result is only 6.3µm away from the screening limit, which would also be considered extremely close to an ideal result indicating good measurement agreement in proximity of the screening limit.

In the GMUTS system, the standard deviation of measurement data is calculated as well and compared with a predefined table of allowable levels and ranked from one to 10. Since no author-known tables exist, repeatability results for this data have not been included here. To get a better evaluation of repeatability, the next section will review the application of the GR&R tool to the set of RRO data.

Tool C application: GR&R

Consider the results below in Table 1 obtained by putting the example data through an ANOVA-based Gauge Reproducibility & Repeatability (GR&R) analysis in Minitab, which is a statistical tool that allows users to analyze numerous types of data (Minitab Company Information, 2010).

Determining the statistical significance of individual factors in an ANOVA analysis typically requires a probability value (P-value) less that 0.05 (5%). That is, if the P-value is less than 0.05 it is considered to be a statistically significant parameter effecting the change of values of the data. Under the Probability (P) column for the results in Table 1, there are two P-values shown to be 0.000. These two values indicate that the measurement values of the individual tires themselves and the interaction between the individual tires and the type of machine are statistically significant. This is expected, since we would hope that the measurement machine is repeatable enough to distinguish the measurements of each tire.

The P-value of the machine factor is 0.295 indicating that it is not known to be a significant factor. In other words, there is no statistically significant difference between the measurements taken on the AkroDYNE Dynamic Balancer and on the ASTEC PLUS Uniformity Machine. By looking at more GR&R results as shown in Table 2, it can be seen that the two different machines contribute only 0.01% of the total variation.

This is quite small and certainly indicates that the difference in measurement machines is not a significant factor that changes the measurement values. One other result that can be calculated is the % Study Variation and % Tolerance Variation. These are calculated and shown in Table 3.

The general Automotive Industry Action Group (AIAG) publishes general guidelines for acceptable limits of both % Study Variation (to assess usefulness of

Table 1: Gage R&R study – ANOVA method for RRO

Two-way ANOVA table with interaction

Source	DF	SS	MS	F	P
Tire No.	14	8145437	581817	223.057	**0.000**
Machine	1	3038	3083	1.182	**0.295**
Tire No. *Machine	14	36517	2608	5.605	**0.000**
Repeatability	120	55840	465		
Total	149	8240877			

Table 2: GR&R % contribution results

Source	VarComp	%Contribution (of VarComp)
Total gage R&R	900.3	1.53
-Repeatability	465.3	0.79
-Reproducibility	434.9	0.74
-Machine	6.3	**0.01**
-Machine*Tire	428.6	0.73
Part-to-part	57920.9	98.47
Total variation	58821.1	100.00

measurement device for process improvement) and % Tolerance Variation (to assess usefulness of measurement device as a screening tool). These guidelines state that values less than 10% are considered excellent, between 10% and 30% is acceptable, and above 30% is unacceptable. By focusing on the Machine term in Table 3, the % Study Variation is 1.04% and the % Tolerance Variation is 1.51%. These values are small when compared to the AIAG excellence guidelines.

One of the advantages of using this technique is that repeatability of the machines is calculated as well. In this example, the repeatability is 8.89% of Study Variation overall and 12.94% of Tolerance overall. Both are acceptable according to AIAG guidelines. The entire process in this example, including the repeatability and measurement agreement elements, result in acceptable Study Variation and Tolerance Variation.

The bottom line

As demonstrated in the previous example, there are several tools that may be used to assess measurement agreement between two or more final finish testing systems. Each of these tools must be examined and monitored in order to ensure they are used correctly to accurately assess machine agreement. Key elements that help assure this process can be classified into three major categories.

First, forcing machine measurement agreement in the absence of measurement machine accuracy is a pivotal flaw when determining whether two or more machines are in agreement. When tire manufacturers choose their measurement agreement standard, it is very important not to adjust new machinery to previously selected master tire values but rather focus on improving the accuracy of the individual machine.

Final finish testing machines need to be accurate in order to ensure measurement agreement in the future. Master tire values should only be used as a guide for the manufacturer when troubleshooting machinery if the accuracy of the machine is not fully known.

Table 3 – GR&R % study and % tolerance variation for process tolerance of 1,000µm

		Study Var	%Study Var	%Tolerance
Source	StdDev (SD)	(6 * SD)	(%SV)	(SV/Toler)
Total gage R&R	30.004	180.03	12.37	18.00
-Repeatability	21.572	129.43	**8.89**	**12.94**
-Reproducibility	20.885	125.13	8.60	12.51
-Machine	2.515	15.09	**1.04**	**1.51**
-Machine*Tire No.	20.703	124.22	8.54	12.42
Part-to-part	240.668	1444.01	99.23	144.40
Total variation	242.531	1455.18	100.00	145.52

Number of distinct categories = 11

Second, there are several factors to consider when choosing the best tool for determining machine agreement:

- The first deals with choosing a set of master tires to match exactly what the manufacturing application requires. These tires will be run repeatedly and it is very important to update master tire values, or after some time, develop new master tires in order to replace worn tires in order to ensure accuracy and agreement for a long period of time.
- The machine agreement assessment tool should devote special attention in the neighborhood of the anticipated screen limits as force machine correlation at extremely high or low measurement values is of secondary importance.
- You must not waste time and energy on machines that manufacturers know are not repeatable. It is inherent that some machines will not offer great repeatability and if they do not, then it will be extremely difficult to analyze measurement agreement on those machines.

Third, for tire manufacturers to be successful in determining machine agreement, it is pivotal for them to work with suppliers that understand these processes and approaches. This may seem to be an obvious decision, but it is very important for manufacturers to follow this point. This may mean the difference between machines that are in agreement (correlation) and testing properly, and machines that are not in agreement, whose time and resources are being misused.

It is clear that these decisions affect more than the manufacturing facility itself. These decisions require sound planning and collaboration with suppliers that know the tools best. Ensuring that tires delivered to the marketplace are what they claim to be is the goal of every tire manufacturer. Verification of such is the function of every final finish operation. Ensuring that tire testing machines read and correlate correctly is a key step to assure the customer, who is demanding a desirable ride on those tires. ***tire***

Bibliography

1) *Anderson, S. A. (2005). Statistics for Business and Economics. Mason: Thomson South Western.*

2) *Correlation (Pearson, Kendall, Spearman). (2009). July 27, 2010, from Statistics Solutions: http://www.statisticssolutions.com/methods-chapter/statistical-tests/correlation-pearson-kendall-spearman/*

3) *Minitab Company Information. (2010). July 27, 2010, from Minitab - Software for Quality Improvement: http://www.minitab.com/en-US/company/default.aspx*