Discussion of "A Risk-Based Measure of Time-Varying Prognostic Discrimination for Survival Models," by C. Jason Liang and Patrick J. Heagerty.[Q1]

H. Michael[1,*] and L. Tian[2,][Q2]**

[1]Department of Statistics, Stanford University, California 94305, U.S.A.
[2]Department of Biomedical Data Science, Stanford University, California 94305, U.S.A.
[*]*email:* haben.michael@stanford.edu
[**]*email:* lutian@stanford.edu

1. Introduction

We congratulate Liang and Heagerty for their clever generalization of the discrimination slope (DS) to allow for evaluation of the time-varying discrimination performance of a biomarker in survival analysis. Lately, there has been growing interest in developing good measures of the performance of a biomarker (or a set of biomarkers) for predicting survival outcomes. The appropriate choice and interpretation of such measures are crucial, for example, in evaluating the incremental value of a novel biomarker such as CRP or coronary artery calcium score relative to conventional risk factors in predicting future cardiovascular risk for an individual patient (Lloyd-Jones et al., 2006). To this end, many generalizations of summary statistics such as $c-$index and R^2 have been proposed. Liang and Heagerty's proposed metric is unique in being based on the risk function, making it particularly suitable for hazard function-based regression models such as the popular Cox model. Specifically, Liang and Heagerty propose the ratio

$$HDS(t) = \frac{\mathrm{E}_{M|T=t}\left\{\lambda(t|M)\right\}}{\mathrm{E}_{M|T>t}\left\{\lambda(t|M)\right\}}$$

to measure the predictive ability of a biomarker M for survival outcome T at time t, where $\lambda(\cdot|m)$ is the hazard function of T conditional on $M = m$. An interesting equivalent formulation of $HDS(t)$ not mentioned in the article is

$$HDS(t) = 1 + \frac{\mathrm{Var}_{M|T>t}\left\{\lambda(t|M)\right\}}{\left[\mathrm{E}_{M|T>t}\left\{\lambda(t|M)\right\}\right]^2},$$

that is, $HDS(t)$ is the 1 plus the squared coefficient of variation of $\lambda(t|M)$ conditional on $T > t$. This form shows that $HDS(t)$ is always ≥ 1 and $= 1$ if and only if $\lambda(t|m)$ does not depend on the marker value m. In the Cox regression model $\lambda(t|m) = \lambda_0(t)\exp(\beta_0 m)$, this form also suggests a simple estimator for $HDS(t)$ based on the empirical mean and variance of $\{\exp(\hat{\beta}' M_j) \mid X_j \geq t\}$, where $X_j = \min(T_j, C_j)$, C_j is the censoring time independent of (M_j, T_j), and $\hat{\beta}$ is the maximum partial likelihood estimator of the regression coefficient β_0.

2. Alternative Generalization of DS

While $HDS(t)$ is an appealing companion to the hazard-based regression model, its interpretation may not be entirely straightforward. A technical difficulty is that the expectation of the conditional hazard function with respect to the marker is not the hazard function of a well defined survival distribution. By contrast, for the risk $\mathrm{P}(D = 1|M = m)$ used in the original definition of DS for binary outcomes, the expectation is simply the population risk when the marker M follows the distribution of interest, that is, the distribution of either cases or controls. More importantly, while the risk is directly connected with and almost sufficient for predicting the binary outcomes D, the hazard function only discriminates between events $T \in [t, t+\epsilon]$ and $T \in (t+\epsilon, \infty)$ for very small $\epsilon > 0$, which is typically not the problem of interest. In addition, it would be sensible to always predict $T \in (t+\epsilon, \infty)$ since $P(T \in [t, t+\epsilon]) \approx 0$. In this sense, $HDS(t)$ at a single time point t is not associated with a meaningful prediction problem, and only collectively may $\{HDS(t), t \in [0, \tau]\}$ be said to reflect the predictive performance of the marker, where τ is some constant smaller than the maximum follow-up time such that $\mathrm{P}(X \geq \tau) > 0$. Even this statement is vague, however, since we have not fully specified what to predict and how to predict it. Furthermore, there is no straightforward way to summarize $\{HDS(t), t \in [0, \tau]\}$ by a single number, such as the area under the ROC curve, which provides practical users with a simple interpretation. To overcome the aforementioned difficulties, we propose to tackle the problem by specifying in order:

(i) The prediction target;
(ii) The prediction method;
(iii) The measurement of quality of prediction.

In practice, one may be interested in predicting whether a patient survives beyond a time point t_0, that is, a binary vari-

able $D = I(T \geq t_0)$. In this simple case, the traditional DS can be used to assess performance of prediction. Alternatively, one may want to predict the survival time itself as a continuous variable. However, due to the presence of right censoring, the distribution of the survival time is often not identifiable on the entire support of T, and it becomes impossible to predict survival time beyond the maximum follow-up time without artificial extrapolation. One practical alternative is to predict the truncated survival time $T \wedge \tau$. A natural prediction for truncated survival time is the conditional expectation

$$\mu(m) = \mathrm{E}(T \wedge \tau | M = m),$$

the mean restricted survival time (RMST). Graphically, RMST represents the area under the survival curve of T given $M = m$ between 0 and τ (Zhang and Schaubel, 2012; Uno et al., 2014; Tian et al., 2014). With both prediction target and method determined, the last step is to generalize DS to reflect the performance of M as a predictor in this setting:

$$\mu(m) \to T \wedge \tau.$$

Recall that for a binary response DS is defined as

$$\mathrm{E}_{M|D=1}\{\mu_D(M)\} - \mathrm{E}_{M|D=0}\{\mu_D(M)\},$$

measuring the separation between the two distributions $M|D = 1$ and $M|D = 0$ by the conditional expectation $\mu_D(m) = \mathrm{P}(D = 1|M = m) = E(D|M = m)$. In the survival analysis setting, it is thus natural to use $\mu(m)$ to assess separation among a family of distributions $M|T = t$ indexed by $t \in [0, \tau]$. While there are many notions of distance between distributions, the difference in the expectation of $\mu(M)$ is directly associated with the performance of the prediction method. Specifically, let us define

$$R(t) = \mathrm{E}_{M|T\wedge\tau=t}\{\mu(M)\}, t \in [0, \tau],$$

so that the slope of the curve $R(t)$ reflects the performance of the marker M at predicting $T \wedge \tau$ by $\mu(M)$.

Furthermore, we may use

$$GDS = \frac{\int_0^\tau R(t)(t - \mu_T)dF_\tau(t)}{\int_0^\tau t(t - \mu_T)dF_\tau(t)}$$

to represent this slope, where $\mu_T = \mathrm{E}(T \wedge \tau)$ and $F_\tau(\cdot)$ is the cumulative distribution function of $T \wedge \tau$. GDS is the least squares estimator of α_1 in fitting the linear model $R(T \wedge t) \sim \alpha_0 + \alpha_1 T \wedge t$ and thus a sensible metric for the slope of $R(t)$. The definition may be seen to be in direct analogy with binary outcome DS by writing the latter as

$$\frac{\sum_{d=0}^1 \mathrm{E}_{M|D=d}\{\mu_D(M)\}(d - p_D)\mathrm{P}(D = d)}{\sum_{d=0}^1 d(d - p_D)\mathrm{P}(D = d)},$$

where $p_D = \mathrm{E}(D)$. Furthermore, after some algebra, GDS may be written as

$$\frac{\mathrm{Var}\{\mu(M)\}}{\mathrm{Var}(T \wedge \tau)},$$

the proportion of variation of $T \wedge \tau$ explained by the biomarker, that is, a straightforward generalization of R^2 from ordinary least squares methods. The justification for this claim is given in the appendix.

In practice, both $R(t)$ and GDS are unknown and need to be estimated from the observed data. Under the proportional hazards model, $R(t)$ can be estimated as

$$\hat{R}(t) = \begin{cases} \frac{\sum_{j=1}^n \hat{\mu}(M_j)\exp\{\hat{\beta}M_j - \hat{\Lambda}_0(t)e^{\hat{\beta}M_j}\}}{\sum_{j=1}^n \exp\{\hat{\beta}M_j - \hat{\Lambda}_0(t)e^{\hat{\beta}M_j}\}}, & t \in [0, \tau) \\ \frac{\sum_{j=1}^n \hat{\mu}(M_j)\exp\{-\hat{\Lambda}_0(t)e^{\hat{\beta}M_j}\}}{\sum_{j=1}^n \exp\{-\hat{\Lambda}_0(t)e^{\hat{\beta}M_j}\}}, & t = \tau \end{cases}$$

where

$$\hat{\mu}(m) = \int_0^\tau \exp\{-\hat{\Lambda}_0(t)e^{\hat{\beta}m}\}dt$$

and $\hat{\Lambda}_0(\cdot)$ is the Breslow estimator for the cumulative baseline hazard function $\Lambda_0(t) = \int_0^t \lambda_0(s)ds$. The estimator of GDS may then be constructed easily from $\hat{R}(t)$.

Although a single biomarker M is used for illustration, the proposal can be conveniently generalized to evaluate the predictive performance of a set of biomarkers. The proposal may also be applied with other functions of M, such as the conditional median, as predictor of the truncated survival time, although the interpretation of GDS as R^2 may be lost.

More generally, the statistic $R(t)$ is non-parametric insofar as it is defined, like $HDS(t)$, independently of any specific regression model. The separation is important since any model is approximate and evaluation metrics tied to a model may be too restrictive. In the current setting, $R(t)$ and GDS can be non-parametrically estimated for any given predictor $\mu(\cdot)$ of the truncated survival time. When $\mu(t)$ is model-based, we may further calibrate it non-parametrically to improve the prediction accuracy and thus increase the associated GDS. On the other hand, modeling may be expedient or even necessary for the purpose of statistical inference. As indicated above, under the proportional hazards model, $HDS(t)$ and $R(t)$ can be estimated conveniently and admit useful statistical inferences, without resorting to relatively complicated techniques like smoothing. Moreover, additional modeling may be unavoidable when, as in the example below, a set of biomarkers is being evaluated, since non-parametric inference in a multi-dimensional setting would require unrealistically large samples. An intermediate approach in this case is first to combine the biomarkers into a single score $\tilde{S}(M)$ via an appropriate regression model and then construct an estimator for $E\left\{T \wedge \tau | \tilde{S}(M)\right\}$ as the predictor for truncated survival time, for which $R(t)$ and GDS can be subsequently evaluated.

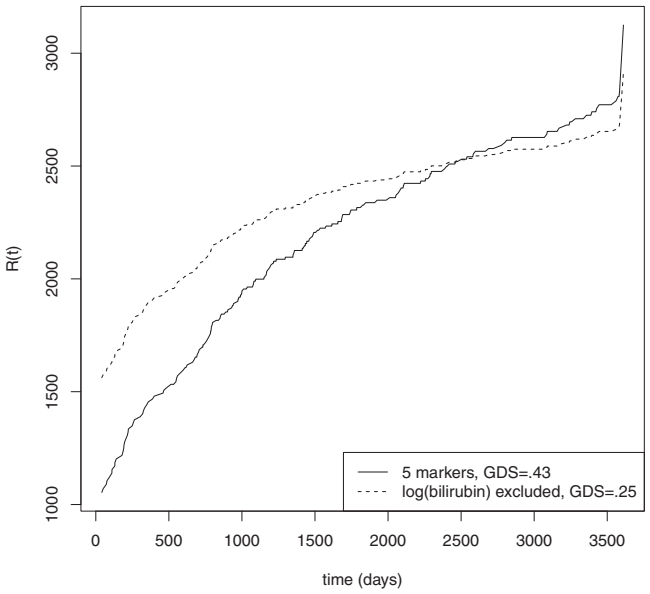


Figure 1. $R(t)$ curve based on (1) log(bilirubin), log(prothrombin time), edema, albumin, and age (solid) and (2) log(prothrombin time), edema, albumin, and age (dotted). The Cox regression model is assumed in estimating $R(t)$ and the associated GDS.

3. Example

Using the Mayo PBC data analyzed by Liang and Heagerty, we estimate $R(t)$ under the Cox regression model. We first consider the five covariates log(bilirubin), log(prothrombin time), edema, albumin, and age, and then perform the analysis again after dropping log(bilirubin). Survival time was truncated at $\tau = 10$ years. The results are presented in Figure 1. While the GDS of the full model is 0.46, it drops to 0.29 after excluding log(bilirubin), known to be an important predictor. Coupled with GDS summaries, the two $R(t)$ curves in Figure 1 provide useful information on ability of the biomarkers to predict truncated survival time with RMST. For example, the slopes of the two $R(t)$ curves are more steep in the first 2–3 years, suggesting that the selected biomarkers are more informative in differentiating survival time among high risk patients. This is consistent with the observations of Liang and Heagerty with regard to $HDS(t)$.

4. Discussion

Liang and Heagerty's HDS statistic is a great leap forward in measuring the performance of a predictor in the survival analysis setting. The important aspect of HDS is that it uses the hazard function $\lambda(\cdot|M)$, which characterizes the distribution of survival outcomes, in comparing the distributions of $M|T = t$ with that of $M|T > t$. However, some difficulties in interpreting the new metric remain. Instead of attempting to propose an intrinsic measure not directly associated with any particular prediction method, it may be more appropriate to choose a metric according to the prediction method of interest. In survival analysis, we may be interested in predicting the truncated survival outcome via RMST, where a more direct generalization of the DS is available. We would be very interested to hear comments from Liang and Heagerty on the issue of coupling the definition of the metric for predictive quality with the concrete prediction method of interest.

Acknowledgements

This work is supported by NIH grants R01 HL089778 and R21 DK103118-01.

References

Lloyd-Jones, D. M., Liu, K., Tian, L., and Greenland, P. (2006). Narrative review: Assessment of C-reactive protein in risk prediction for cardiovascular disease. *Annals of Internal Medicine* **145**, 35–42.

Zhang, M. and Schaubel, D. (2012). Double-robust semiparametric estimator for differences in restricted mean lifetimes in observational studies. *Biometrics* **68**, 999–1009.

Tian, L., Zhao, L., and Wei, L. J. (2014). Predicting the restricted mean event time with the subject's baseline covariates in survival analysis. *Biostatistics* **15**, 222–233.

Uno, H., Claggett, B., Tian, L., Inoue, E., Gallo, P., Miyata, T., et al. (2014). Moving beyond the hazard ratio in quantifying the between-group difference in survival analysis. *Journal of Clinical Oncology* **32**, 2380–2385.

Appendix

The Interpretation of GDS as a Coefficient of Determination

We assume that T has continuous conditional and marginal the density functions $f(t|m)$ and $\bar{f}_T(t)$, with corresponding survival functions $S(t|m)$ and $\bar{S}_T(t)$. By definition,

$$\mu(m) = \int_0^\tau sf(s|m)ds + \tau S(\tau|m).$$

Therefore,

$$\begin{aligned}
&\int_0^\tau tR(t)dF_\tau(t) \\
&= \int_0^\tau t \int \mu(m)\frac{f(t|m)f_M(m)}{\bar{f}_T(t)}dm\,\bar{f}_T(t)dt \\
&+ \bar{S}_T(\tau)\tau \int \mu(m)\frac{S(\tau|m)f_M(m)}{\bar{S}_T(\tau)}dm \\
&= \int \mu(m)\left\{\int_0^\tau tf(t|m)dt + \tau S(\tau|m)\right\} f_M(m)dm \\
&= \int \mu(m)^2 f_M(m)dm = \mathrm{E}\{\mu(M)^2\},
\end{aligned}$$

where $f_M(\cdot)$ is the density function of M. Similarly, $\int_0^\tau R(t)dF_\tau(t) = \mu_T$ (i.e., $E\{R(T \wedge \tau)\} = \mu_T$) and thus

$$\int_0^\tau R(t)(t - \mu_T)dF_\tau(t) = \mathrm{E}\{\mu(M)^2\} - \mu_T^2 = \mathrm{Var}\{\mu(M)\}.$$

Coupled with the fact that $\int_0^\tau t(t-\mu_T)dF_\tau(t) = \mathrm{Var}(T\wedge\tau)$, this implies that

$$GDS = \frac{\mathrm{Var}\left\{\mu(M)\right\}}{\mathrm{Var}(T\wedge\tau)}.$$

It is clear that GDS also equals

$$\frac{\mathrm{Cov}\left\{\mu(M), T\wedge\tau\right\}}{\mathrm{Var}(T\wedge\tau)},$$

which is the minimizer of the L_2 loss

$$l(\alpha) = \mathrm{E}\left[\left\{R(T\wedge\tau)-\mu_T\right\} - \alpha(T\wedge t-\mu_T)\right]^2.$$

Therefore, GDS also can be viewed as the coefficient of $T\wedge t$ in the linear regression model

$$R(T\wedge t) \sim \alpha_0 + \alpha_1 T\wedge t.$$

AUTHOR QUERY FORM

JOURNAL: BIOMETRICS

Article: BIOM12631

Dear Author,
During the copyediting of your paper, the following queries arose. Please respond to these by annotating your proofs with the necessary changes/additions using the E-annotation guidelines attached after the last page of this article.

We recommend that you provide additional clarification of answers to queries by entering your answers on the query sheet, in addition to the text mark-up.

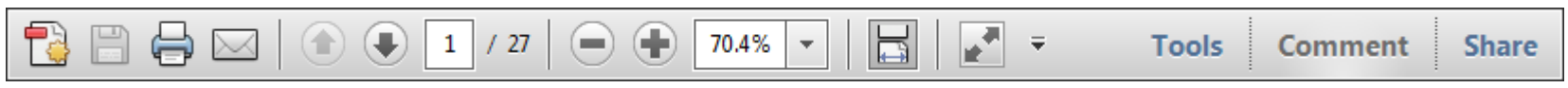
Query No.	*Query*	*Remak*
Q1:	Please check the font used in inline and display equations and confirm that its appearance is correct.	Fine as is
Q2:	Please confirm that given names (red) and surnames/family names (green) have been identified correctly.	Fine as is

USING e-ANNOTATION TOOLS FOR ELECTRONIC PROOF CORRECTION

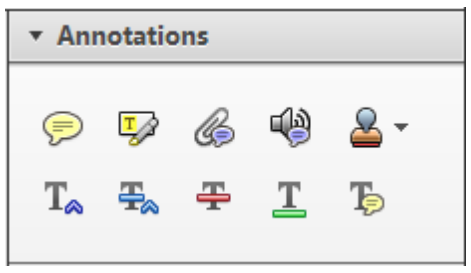
WILEY

Required software to e-Annotate PDFs: Adobe Acrobat Professional or Adobe Reader (version 8.0 or above). (Note that this document uses screenshots from Adobe Reader X)
The latest version of Acrobat Reader can be downloaded for free at: http://get.adobe.com/reader/

Once you have Acrobat Reader open on your computer, click on the Comment tab at the right of the toolbar:



This will open up a panel down the right side of the document. The majority of tools you will use for annotating your proof will be in the Annotations section, pictured opposite. We've picked out some of these tools below:

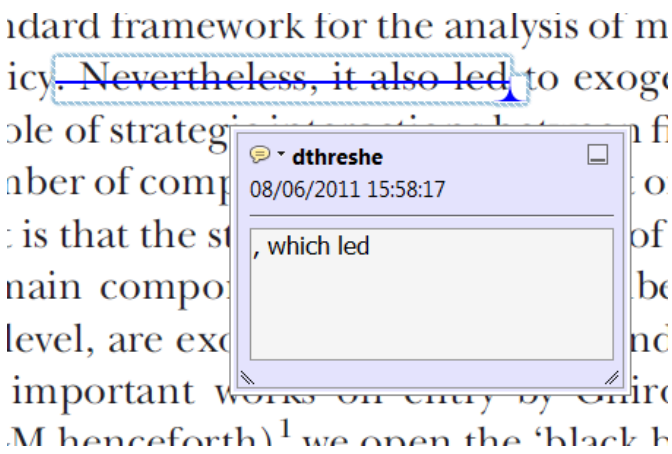


1. Replace (Ins) Tool – for replacing text.

Strikes a line through text and opens up a text box where replacement text can be entered.

How to use it

- Highlight a word or sentence.
- Click on the Replace (Ins) icon in the Annotations section.
- Type the replacement text into the blue box that appears.

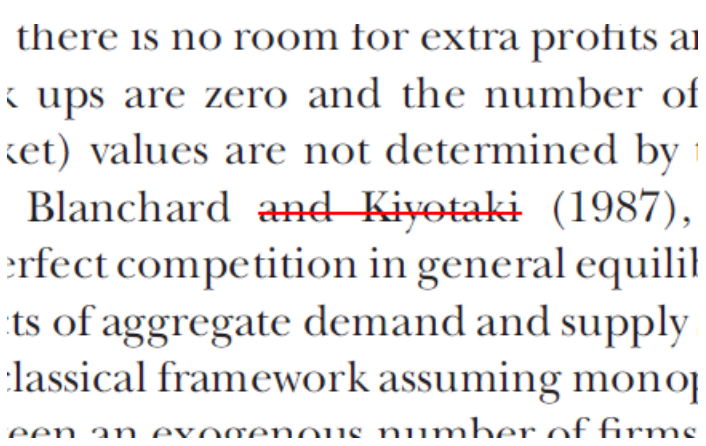


2. Strikethrough (Del) Tool – for deleting text.

Strikes a red line through text that is to be deleted.

How to use it

- Highlight a word or sentence.
- Click on the Strikethrough (Del) icon in the Annotations section.

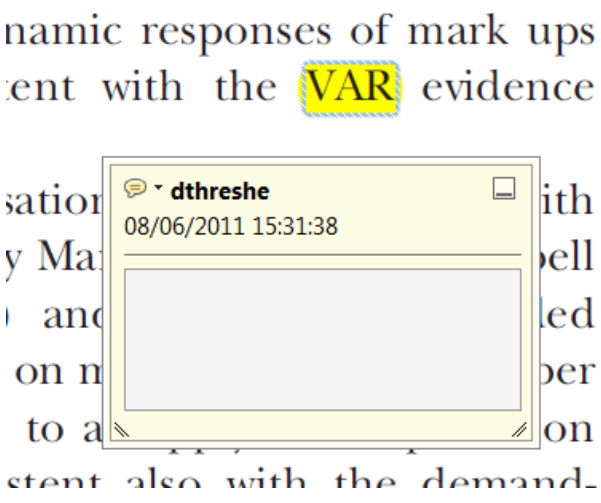


3. Add note to text Tool – for highlighting a section to be changed to bold or italic.

Highlights text in yellow and opens up a text box where comments can be entered.

How to use it

- Highlight the relevant section of text.
- Click on the Add note to text icon in the Annotations section.
- Type instruction on what should be changed regarding the text into the yellow box that appears.

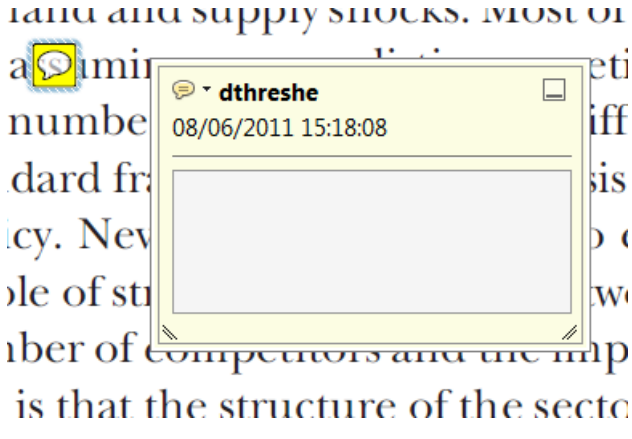


4. Add sticky note Tool – for making notes at specific points in the text.

Marks a point in the proof where a comment needs to be highlighted.

How to use it

- Click on the Add sticky note icon in the Annotations section.
- Click at the point in the proof where the comment should be inserted.
- Type the comment into the yellow box that appears.



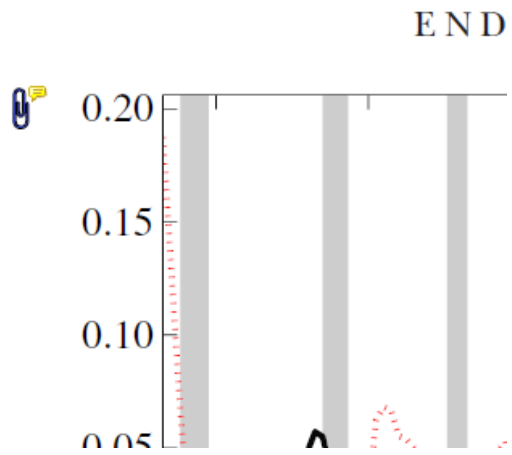
USING e-ANNOTATION TOOLS FOR ELECTRONIC PROOF CORRECTION

5. Attach File Tool – for inserting large amounts of text or replacement figures.

Inserts an icon linking to the attached file in the appropriate pace in the text.

How to use it

- Click on the Attach File icon in the Annotations section.
- Click on the proof to where you'd like the attached file to be linked.
- Select the file to be attached from your computer or network.
- Select the colour and type of icon that will appear in the proof. Click OK.

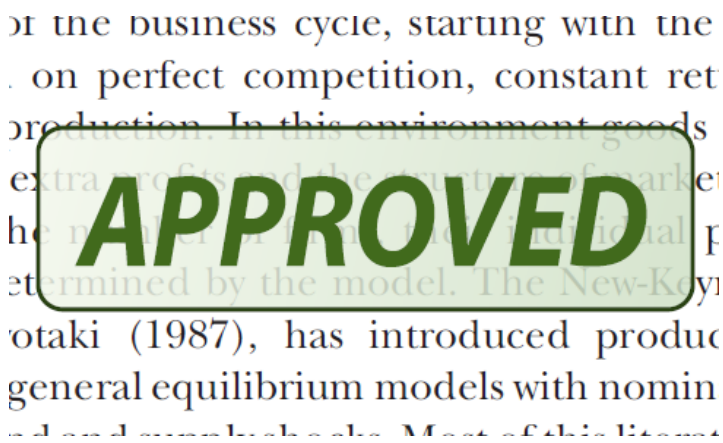


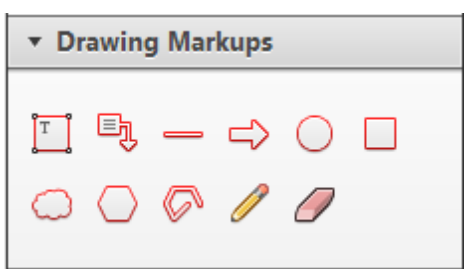
6. Add stamp Tool – for approving a proof if no corrections are required.

Inserts a selected stamp onto an appropriate place in the proof.

How to use it

- Click on the Add stamp icon in the Annotations section.
- Select the stamp you want to use. (The Approved stamp is usually available directly in the menu that appears).
- Click on the proof where you'd like the stamp to appear. (Where a proof is to be approved as it is, this would normally be on the first page).



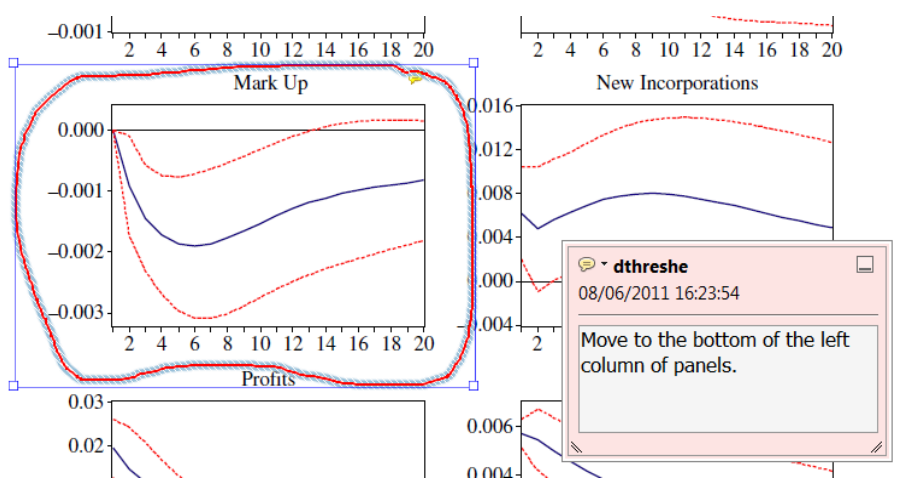


7. Drawing Markups Tools – for drawing shapes, lines and freeform annotations on proofs and commenting on these marks.

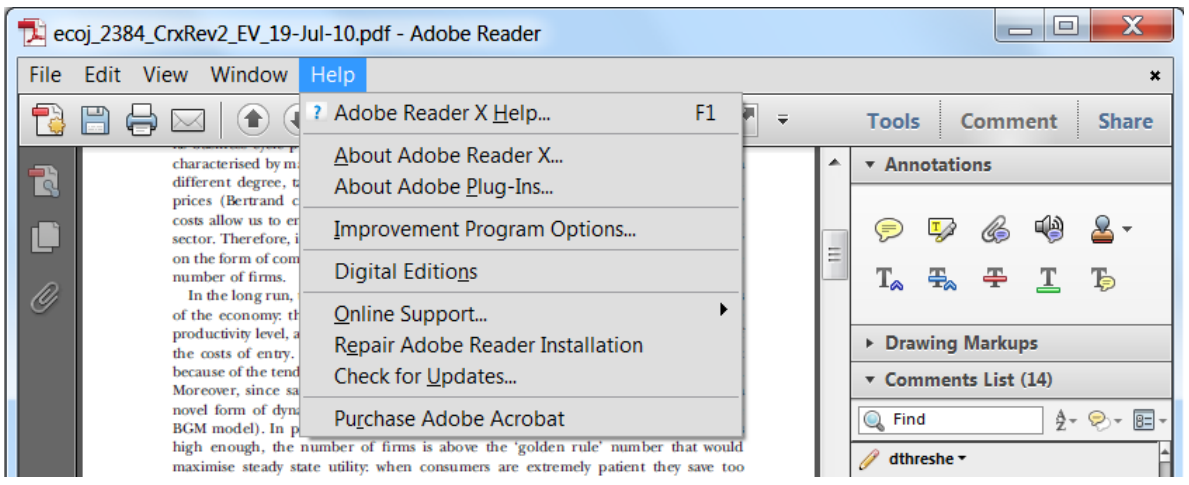
Allows shapes, lines and freeform annotations to be drawn on proofs and for comment to be made on these marks..

How to use it

- Click on one of the shapes in the Drawing Markups section.
- Click on the proof at the relevant point and draw the selected shape with the cursor.
- To add a comment to the drawn shape, move the cursor over the shape until an arrowhead appears.
- Double click on the shape and type any text in the red box that appears.



For further information on how to annotate proofs, click on the Help menu to reveal a list of further options:



Proof Correction Marks

Please correct and return your proofs using the proof correction marks below. For a more detailed look at using these marks please reference the most recent edition of The Chicago Manual of Style and visit them on the Web at: http://www.chicagomanualofstyle.org/home.html

Instruction to typesetter	*Textual mark*	*Marginal mark*
Leave unchanged	under matter to remain	stet
Insert in text the matter indicated in the margin	∧	∧ followed by new matter
Delete	through single character, rule or underline or through all characters to be deleted	
Substitute character or substitute part of one or more word(s)	through letter or through characters	new character or new characters
Change to italics	under matter to be changed	ital
Change to capitals	under matter to be changed	caps
Change to small capitals	under matter to be changed	sc
Change to bold type	under matter to be changed	bf
Change to bold italic	under matter to be changed	bf+ital
Change to lower case		lc
Insert superscript	∨	∨ under character e.g. ∨
Insert subscript	∧	∧ over character e.g. ∧
Insert full stop	⊙	⊙
Insert comma		
Insert single quotation marks		
Insert double quotation marks		
Insert hyphen		
Start new paragraph		
Transpose		
Close up	linking characters	
Insert or substitute space between characters or words	#	#
Reduce space between characters or words		