Using a frailty model to measure the effect of covariates on the disposition effect

Julia Brettschneider and Matthew Burgess

Paper No. 17-05, www.warwick.ac.uk/go/crism

Abstract

Proportional hazards regression models have previously been used to measure the effect of covariates on the disposition effect (DE) in datasets of individual investor trading records. Correlation between trades made by the same investor can be problematic for models of this kind. Past approaches have used a marginal model, where standard errors robust to investor-level correlation are computed after the model has been estimated. This report explores the use of frailty models as an alternative, where the correlation is modelled explicitly through the use of latent variables. Using a dataset of trading records from a large discount brokerage, a frailty model is shown to provide a significantly improved fit relative to the corresponding marginal model. The frailty model is able to significantly estimate a greater number of effects, and it adheres more closely to the important proportional hazards assumption. Results from the frailty model provide some new evidence on experience and learning; the number of trades an investor has made does not have a significant effect on the DE when the investor's self-assessed experience level is included in the model, and the length of time an investor has held an account does not appear to be a reliable measure of experience in this dataset, as those who opened an account most recently exhibit the weakest DE. Graphical checking of the proportional hazards assumption adds nuance to the interpretation of some variables. For example, the weakening of the DE in December is much larger for positions that have already been held for a long period of time, and differences in the DE between positions in small and large cap-size stocks only start to materialize after they have been held for roughly 100 days.

1 Introduction

The disposition effect (DE) is the tendency of investors to sell assets that have increased in price since purchase (gains), and hold assets that have decreased in price (losses). It has been detected in a large number of datasets, particularly the trading records of individual investors.[1] An important aim in this literature has been to determine which factors either strengthen or weaken the DE, particularly the characteristics of the investors themselves. Starting with Feng and Seasholes [2005], the use of a proportional hazards regression model, a method from survival analysis, has been a common approach to this problem. In the context of survival analysis, an important property of trading record datasets is the natural grouping of trades at the investor level. Positions held by the same investor are likely to be dependent due the particular investment style of the investor, and this unobserved heterogeneity needs to be addressed when estimating a regression model. The problem of investor-level dependency is made more important by the extreme imbalance that can be present in datasets of this kind. A small number of investors account for a large proportion of overall trading activity, hence effect estimates will disproportionately reflect the idiosyncrasies of these investors if no effort is made to control for them.

In the existing literature the marginal method is used, which adjusts standard errors after the model has been estimated to account for the investor-level correlation between positions. This report will explore the use of frailty models as an alternative, using a dataset consisting of the trading records of 78,000 investors at a large discount brokerage in the U.S. over the period 1991-96 as well as a variety of demographic information. This dataset was first analysed in Barber and Odean [2000] and has been used extensively in the behavioural finance literature.[2]

In a frailty model, the propensity to sell a position that is unique to each investor

[1] See for example, Odean [1998], Grinblatt and Keloharju [2001], Feng and Seasholes [2005] and Brown et al. [2006].

[2] We are grateful to Terrance Odean for sharing this dataset with us.

is modelled directly in a way that is analogous to the use of random effects in linear regression. For the dataset used here, the addition of a frailty component significantly improves upon the equivalent marginal model and allows a greater number of effects to be significantly estimated. In addition, the frailty model adheres much more closely to the proportional hazards assumption, which makes the parameter estimates more useful as summaries of the effect over time. Explicit checking of the proportional hazards assumption is typically neglected in the existing literature. The analysis below shows that doing so can add important caveats to the interpretation of regression results, with some effects being much stronger or weaker as the holding period of a position increases.

The remainder of the report is organised as follows. Section 2 discusses the problem of investor-level correlation in detail, along with both the marginal and frailty approaches to dealing with with. Sections 3 and 4 describe the dataset and covariates that will be used for model estimation. Section 5 describes how the sample is formed from trading records for estimating a survival model. Section 6 presents results from estimating both a frailty and marginal model, and highlights how they differ. Section 7 presents some additional models that check the robustness of the main results. Section 8 discusses the advantages that using a frailty model provided for this analysis and compares the results to those in the literature.

2 The problem of correlation

Datasets of brokerage trading records are naturally grouped at the investor level. In models where the observational units are stock positions, the possibility of dependence between positions held by the same investor needs to be addressed. Survival models for the lifetimes of stock positions are a popular method for measuring the disposition effect (DE) and factors that may affect its severity, hence investor-level correlation is an important issue in empirical work on the DE. In the existing literature, the marginal model approach has been used, where standard errors that are robust to correlation between positions are calculated after the model has been estimated.[3]

This report will explore the use of frailty models as an alternative, where investor-level correlation is modelled explicitly through the use of latent variables. This method is analogous to the use of random effects in linear regression. Frailty models make assumptions about the structure of dependence within and between groups of positions, whereas no such assumptions are made in the marginal approach. However if these assumptions are reasonable then using a frailty model will provide a number of advantages over the corresponding marginal model. This section will first review the standard Cox model before describing the different modifications that are made to it in marginal and frailty models. The differences between the two methods in practice will then be discussed.[4]

2.1 The Cox model

Let $\lambda(t, X(t))$, for $t > 0$, be the rate at which stock positions are sold, which shall be referred to as the hazard rate. It is dependent on the time t and the p-dimensional vector $X(t)$, which consists of observations on p covariates at time t. The hazard rate, or function, is defined as

$$\lambda(t, X(t)) = \lim_{h \downarrow 0} \frac{\mathbb{P}(t \leqslant T < t + h | T \geqslant t, \{X(u); u \leqslant t\})}{h}$$

where T is the time at which the stock is sold. The hazard function is therefore approximately the probability of the stock being sold in the infinitesimal period of time immediately after t, conditional on the stock not having been sold prior to t and on the 'history' of the covariates during the period $[0, t]$.

[3] Marginal models encompass a wider set of strategies for dealing with complex survival data, but for models of stock position lifetimes, the marginal approach only requires the calculation of robust standard errors.

[4] For a more detailed discussion of the theoretical and practical differences between marginal and frailty models, see Wienke [2010].

Proportional hazards (PH) models assume that the hazard rate can be split into two components. The first, called the baseline hazard function, is common to all positions and depends only on time. The second depends on the current values of the covariates, which may themselves depend on time. The key assumption of PH models is that changes in the covariates have a multiplicative effect on the baseline hazard, and that this effect is constant over time. Checking this assumption will be an important part of the analysis presented below. In such models, the hazard function can be written as

$$\lambda(t, X(t), \beta) = \lambda_0(t) \exp(\beta^\top X(t)) \tag{1}$$

where $\lambda_0(t)$ is the baseline hazard function, and β is a p-dimensional vector of coefficients corresponding to the covariates. Fully parametric PH models specify a form for the baseline hazard function and estimate its parameters along with β. However, there may not be a good basis for choosing one particular form over another. Cox et al. [1972] proposed a method for estimating β whilst leaving the baseline hazard function unspecified. Since the effect of covariates is of primary interest and there is no strong reason for choosing a particular form for the baseline hazard, the Cox model will be used throughout this report. Estimation of the Cox model involves maximizing the 'partial likelihood' that depends only on β, given by

$$l_p(\beta) = \prod_{i=1}^{n} \left(\frac{exp[\beta^\top X_i(T_i)]}{\sum_{j \in R(T_i)} exp[\beta^\top X_j(T_i)]} \right)^{c_i} \tag{2}$$

where n is the number of positions across all investors, and c_i and $R(T_i)$ are the censoring indicator and risk set - concepts from survival analysis that will be explained in the following. Positions are under observation for a set period of time after being purchased, called the follow-up time. If a position has not been sold by the end of its follow-up time, then it is said to be censored. A position is also censored if it has not been sold by the end of the data period. In (2), T_i is the time that position i is sold or censored. The censoring indicator c_i equals one if the position is sold at time T_i and zero if it is censored. $R(T_i)$ is the risk set, containing all positions that have not been sold or censored prior to time T_i, other than position i itself. Hence censored positions only contribute to the likelihood by their presence in risk sets at times when other positions are sold.[5]

The follow up time used in the analysis below is 500 days. This decision was made because for longer holding periods it becomes less likely that the investor is actively considering selling the position, and the position is therefore not informative about the effect of covariates on the decision to sell. Related to this is the time scale that is being used to record a position's holding period and how this impacts the risk set that is constructed when the position is sold or censored. The holding period is defined as the time in days since the stock was first purchased. As a consequence of this, the risk set of positions contributing to the denominator in (2) are those that had been held for the same amount of time as the position which was sold. Hence positions in the risk set did not necessarily exist at the same calendar time as the sold stock, but did exist at the same 'follow-up time'. This means that calendar effects, such as whether the hazard of selling is greater in December months, can be tested through the inclusion of dummy variables in the model.

The log of the likelihood in (2) is twice-differentiable and can therefore be maximised using the Newton-Raphson algorithm. Standard errors for the parameter estimates are obtained as the inverse of the information matrix, which is minus the second derivative of the likelihood for β.

[5]This formula ignores the possibility of tied event times, which occur in this dataset when two or more positions are sold on the same day. The method proposed in Efron [1977] is used to approximate the likelihood contribution for a time point when there are multiple events.

2.2 Marginal model

In the marginal approach, all positions are assumed to be independent for the purpose of estimating β. Estimation proceeds in the same way as the standard Cox model, with robust standard errors being calculated afterwards. Lin and Wei [1989] derive an estimator for the variance of $\hat{\beta}$ in the familiar 'sandwich' form that is consistent in the presence of correlation between groups of units, in this case correlation between positions held by the same investor. Lipsitz et al. [1996] provide an estimator that is asymptotically equivalent to that of Lin and Wei, but easier to compute in practice using an infinitesimal jackknife method.

After a Cox model has been estimated as described above, positions are split into q groups, one for each investor. The jackknife residual for group j is

$$J_j = \hat{\beta} - \hat{\beta}_{-j}$$

where $\hat{\beta}_{-j}$ is the vector of parameter estimates resulting from fitting the same model but after deleting the observations in group j. Fitting a new model for every group would be very costly, so instead $\tilde{\beta}_{-j}$, an approximation of $\hat{\beta}_{-j}$, is computed in the following way. Estimate $\hat{\beta}$ by letting the Newton-Raphson algorithm run until convergence as usual, then run it for another step after deleting the observations in group j, producing $\tilde{\beta}_{-j}$. The approximate jackknife variance estimate is then

$$\left(\frac{q-p}{q}\right)\sum_{j=1}^{q}(\tilde{\beta}_{-j} - \hat{\beta})(\tilde{\beta}_{-j} - \hat{\beta})'$$

where p is the number of parameters being estimated. In the model estimated in section 6 the robust standard errors are significantly larger than their naive counterparts, but in general they can be larger or smaller.

2.3 Frailty model

In contrast to a marginal model, frailty models assume a specific structure for the dependence between positions held by the same investor and incorporate it in the model directly. Specifically, a shared frailty model can be used where each investor is assumed to have a certain propensity for selling positions they hold. Conditional on this shared frailty and the model covariates, the survival times of positions held by the same investor are assumed to be independent. Similarly, the effect of covariates on the survival of positions held by different investors is assumed to be constant, conditional on the investor-specific frailties.

These investor-level frailties are included in the Cox model as fixed quantities that act multiplicatively on the baseline hazard function. In a shared frailty model for position lifetimes with $j = 1, ..., q$ investors and $i = 1, ..., n_j$ positions associated with each investor, the hazard function for the i-th position of the j-th investor is

$$\lambda_{ij}(t) = Z_j \lambda_0(t) \exp(\beta^\top X_{ij}(t))$$

Where X_{ij} is the covariate vector for the position and Z_j the unobserved frailty associated with investor j. The vector of frailties Z are assumed to be random variables with a common distribution, the parameters of which can be estimated. The gamma distribution is a common choice of frailty distribution, and is the one that will be used here. As discussed in Wienke [2010], the gamma distribution is used as a frailty distribution primarily because of its mathematical properties: it produces non-negative values, it can flexibly model a variety of distributional 'shapes' as its variance changes and the simplicity of its Laplace transform means the frailty terms can be easily integrated out of the log partial likelihood in the Cox model.

Since a scaling factor common to the frailties of all subjects can be absorbed into the baseline hazard of the Cox model, the gamma distribution for the frailties is taken to have expectation equal to one. Hence only the variance of the frailty distribution must be estimated as an additional parameter. Testing whether this variance is significantly different from zero is the primary way of establishing if investors do indeed have differing levels of frailty when it comes to selling positions they hold.

Derivations of the partial log-likelihood for a Cox model with gamma frailties are provided by Klein [1992] and Nielsen et al. [1992]. As with many latent variable problems the EM algorithm can be used in this situation, as described by these references. In the R 'survival' package[6], which is used here, the problem is placed in the framework of penalised regression. When gamma frailties are used, this approach produces the same estimates as the EM algorithm, as shown in Therneau and Grambsch [2000]. The authors report from experience that this implementation typically converges significantly faster than an equivalent EM implementation. Details of the estimation procedure can also be found in this reference, and in the documentation accompanying the survival package.

Shared frailty is a natural fit for investor trading data since there are a large number of investors and variation at the investor level is more of a nuisance factor that could obscure the effects of covariates in the model, which are the main focus of the analysis. Grouping variables can also be included as fixed effects, for example gender divides the data into two groups and is included in the models estimated later in the report. But this would not be feasible for investors since there are thousands of investors and many have as little as one position in the dataset. Additionally, there is no need to test hypotheses about differences between specific investors, it is sufficient to have an estimate of the variation between investors in general, as is provided in the frailty model.

Possible reasons for there being such a difference between investors include their preference for risk, their beliefs about the market (e.g. whether there is price momentum or not), their investment objectives and the particular strategy they are

[6]See Therneau [2014] for some detail.

following. For example, some more serious investors may trade very frequently and follow a strategy based on short term changes in stock prices. The holding periods of these investors will therefore be shorter than other investors in the sample. Shared frailty can separate out this kind of difference from the effects of covariates included in the model, that are in theory common across all investors.

2.4 Comparison

Because of the assumptions each model makes, the interpretation of parameter estimates is different in marginal and frailty models (note that any reference to a frailty model will from now on mean specifically a shared frailty model). Marginal models estimate average effects at the population level, with the average being across all investors. Hence the coefficient of a covariate describes the expected difference in survival for positions that differ in this covariate, regardless of which investors hold the positions. In a frailty model, a coefficient describes the difference in hazard between two positions that differ in the covariate, conditional on the frailties of the investors.

Since the marginal model is estimating effects averaged across all investors, the estimates can be biased if the sample either contains positions from only a small number of investors, or the distribution of positions across investors in the sample is very imbalanced. The former problem is not relevant in this dataset since over 5,000 investors are represented in the sample used for estimating the models in section 6.[7] The latter is more of a concern though since the sample is extremely imbalanced in terms of positions per investor. The median number of positions per investor is 7, whilst the maximum is 451 and investors in the top decile of this distribution account for 46% of the total number of positions. If the survival of positions is correlated at the investor level, then coefficient estimates that ignore this, as in a marginal model, will disproportionately reflect the behaviour of

[7] A small number of groups would suggest that the grouping variable should be included using fixed effects, rather than as a frailty term.

these most active investors. By separating the coefficient estimates out from each investor's static propensity for selling positions, the frailty model can provide a more accurate summary of the effect a covariate has on the hazard of selling.

Another advantage of frailty models is their ability to explain apparent violations of the PH assumption when effects appear to weaken over the course of the follow-up time. Positions that remain unsold for a long time are more likely to be held by investors with a lower propensity for selling i.e. frailty, hence the effect of all covariates seems weaker when differing levels of frailty are not accounted for in the model. This is more apparent when the follow-up time is long since there is more scope for observing longer survival times. The follow up time here is 500 days, compared to the median holding period of 86 days for positions where a sale is observed, so this will be a relevant issue. Effects that weaken over time can be detected using a graphical method for assessing the proportional hazards assumption, as will be done in section 6 and those following it. The results of section 6 show that for this dataset, adding frailty terms greatly improves the model's adherence to the PH assumption.

If shared frailty is a good description of the dependence structure in the data, then adding a frailty component will produce a more useful model for the reasons described above. In particular it will help isolate the effects of covariates in the model and their interactions, which is the main aim in analyses of the disposition effect. Assessing whether the addition of frailties does improve the model will be the focus of section 6.

3 Dataset

The dataset that will be used to estimate the models below consists of the trading records for 78,000 investors in the U.S. during the period from January 1991 to November 1996. The data were obtained by Odean from a large discount brokerage and are commonly referred to in the literature as the LDB dataset. A detailed description of the dataset can be found in Barber and Odean [2000]. The LDB dataset is notable for the large number of investors it contains, the relatively long period of time it covers and the range of demographic information about the investors that was collected alongside the trading records.

As reported in Goetzmann and Kumar [2008], 62,387 investors trade common stocks, which is the only asset type considered here. The median investor holds a portfolio consisting of three stocks, with a total value of $13,869. Barber and Odean [2000] find that the mean investor in this dataset turns over 75% of their stock portfolio each year and underperforms the market by 1.5% annually. More details about the LDB dataset can be found in the latter reference. Data on stock prices, SIC industry and capitalization were obtained from the CRSP.

4 Covariates

This section contains a description of the covariates that will be included in the model. Their effect on the DE will also be tested through the use of interaction terms. Following previous work on the topic, primarily Feng and Seasholes [2005] (F&S) and Dhar and Zhu [2006] (D&Z), many of them are related to the hypothesis that more sophisticated and experienced investors suffer less from the DE. Some stock-level variables will also be tested. Since the dataset is large, it should be possible to significantly estimate even small effects.

- Paper status: this variable records whether the position is currently trading at a gain (daily low and high prices above the purchase price), loss (daily low and high prices below the average purchase price) or neither. This third state shall be referred to as paper neutral. If positions in the paper gain group have a greater hazard of being sold than those in the paper loss category, then there is a DE.
- Gender: coded as an indicator which equals one if the investor is male. 92% of investors are male, but significant differences between genders have been

found in the LDB dataset. Barber and Odean [2001] find that men trade 45% more than women and that their net return is roughly one percentage point lower per year as a result. However F&S note that, in a different dataset, gender only appears important in explaining the propensity to sell when few other control variables are included. Since a number of controls are being included here, further evidence on the importance of gender can be found.

- Age: the investor's age, as recorded on June 8 1997, 7 months after the end of the data period.

- Professional occupation indicator: whether the investor works in a professional occupation or not. The category is labelled as 'professional/technical', and contains 42% of investors. The remaining investors are in categories such as 'administrative/managerial', 'sales/service', 'clerical/white collar' or are retired. D&Z find that investors who work in professional occupations exhibit a disposition effect that is 20% weaker than those who work in non-professional occupations.

- Income: provided by the brokerage house is a proxy for income that is not continuous yet contains a large number of categories. This variable was therefore converted to a categorical variable by splitting it into quartiles. The four groups shall be referred to as the 'low', 'medium', 'high' and 'very high' income groups. D&Z find that high-income investors exhibit a DE 10% smaller than low income investors.

- Self-assessed experience: When opening an account, investors were asked to describe their investing experience on a four point scale. The levels are 'None', 'Limited', 'Good', and 'Extensive'.

- Tenure: the time in (possibly fractional) years since the investor opened their first account at the brokerage, recorded at the start of the data period. Negative values occur when an investor opened their first account after the start of the data period, as is the case for 12% of investors. Tenure is included as a measure of 'passive' experience, with the theory being that more experience investors will suffer less from the DE. The variable is split into four groups based on the bins defined by (-1, 1], (1, 4], (4, 8] and (8, 16]. Note that the maximum recorded value is 15.8 years. These groups will be referred to as Tenure 1-4, in the same order.

- Initial Diversification: The number of stocks the investor holds, across all their accounts, in the first month for which they have a record in the dataset. This variable is split into bins defined by (0,3], (3,7], (7,12] and (12, ∞), and the groups will be referred to as Diversifictaion 1-4. F&S find that investors with greater initial portfolio diversification suffer less from the DE.

- Number of sales made to date: a time-varying covariate that only includes sales where the purchase occurred after the start of the data period. This variable is highly skewed, so its natural logarithm will be used when estimating models, as has been done in past research. F&S and D&Z find that having completed more sales reduces an investor's DE, implying that investors learn from their mistakes.

- Capitalization quintile of the stock: this data is taken from CRSP and is a time-varying covariate since it is updated annually. Quintile 1 contains stocks with the smallest cap-size, and quintile 5 those with the largest. Exploratory analysis found that positions in stocks from higher cap-size quintiles exhibited a lower DE. Larger cap-size stocks have lower volatility, hence this finding is consistent with Kumar [2009], who finds that lower volatility stocks exhibit less of a DE. For each cap-size quintile, table 4 gives the percentage of holding days across all positions for which the stock held was in that quintile. Quintile 5 was used as the reference level for this variable in the models below, to increase the chance of finding significant differences between groups.

Cap-size quintile	1	2	3	4	5
% of holding days	2.29	4.77	8.14	14.45	70.35

- SIC major group: also from CRSP, this variable separates stocks into one of 10 industry groups. These are described in table 4, which also gives the percentage of positions due to stocks in each of the groups. This variable is included mainly as a control for possible correlation between positions in similar stocks. The manufacturing group (MAN) was used as the reference level for this variable.

SIC major group	group label	% of positions
Agriculture, forestry & fishing	AFF	0.03
Construction	CON	0.57
Finance, insurance & real-estate	FIRE	8.52
Manufacturing	MAN	48.64
Mining	MIN	3.81
Public administration	PUBA	0.13
Retail trade	RETL	7.82
Services	SERV	12.29
Transport & public utility	TRAN	9.75
Wholesale trade	WHOL	2.46

- December indicator: equalling one for intervals of the position's holding period which are in December, and is hence time-varying. Due to the ending of the tax year, trading behaviour is significantly different in December. Odean [1998] finds that the DE is actually reversed for December months.

- 1995/96 indicator: equalling one for intervals of the position's holding period which are in years 1995 or 1996. Exploratory analysis showed that the aggregate DE decreases significantly in 1995/96, the last two years of the data period. Controlling for this effect will therefore be important when trying to isolate the effect of other variables.

5 Sample formation

Positions are defined and recorded as follows. A position starts when an investor purchases a stock they do not already hold. There is information on the stocks an investor holds at the start of the data period, but not the purchase date or price, hence these pre-existing positions cannot be included, and are ignored for the purpose of determining if a stock purchase marks the start of a new position or not. If an investor purchases more of a stock they already hold, the quantity held and share-weighted average purchase price are updated accordingly. A position ends when the stock is sold. If only part of the quantity held is sold, the position is still considered to have ended on that date and will be entered into the dataset as such. This is because the holding period and reference price (the price used to determine if a position is trading at a gain or not, and if a sale was for a gain or not) after the first sale are ambiguous.

If a partial sale has occurred, the quantity held is still tracked and a new position in this stock cannot start until after the entire quantity held has been sold. There are some cases where a sale is made for a greater quantity of stock than was observed being purchased, due to the investor already holding some of the stock before the start of the data period. In these cases the sale is not considered valid for the purposes of this dataset and the quantity held is set to zero. A subsequent purchase of the stock will start a new position as usual.

A position can also end by being censored. A follow-up time of 500 days was chosen, since for much longer holding periods it becomes less likely that the investor is actively considering selling the position, and the position is therefore not informative about the effect of covariates on the decision to sell. The choice of 500 days is somewhat arbitrary, but the decrease in frequency of sales up to and beyond this point is smooth so it will not have a large impact on the results of the analysis. A position can also be censored if a sale has not been observed before the end of the data period, which is November 29th, 1996. In the LDB dataset there are approximately 618,500 valid positions according to the criteria set out above. They are held by 55,300 unique investors and 70% are sold before being censored, for a total of 436,100 sales.

On each day during the holding period of a position, including the day it is sold or censored but not the day it is purchased, the paper status of the position is recorded. Following Odean [1998] The position is trading at a gain if the daily low price is above the average purchase price, at a loss if the daily high price is below the average purchase price, and is considered 'neutral' otherwise.[8]

For fitting the models described in the following sections, the sample was restricted to only those investors and positions for which full information was available on all covariates described in the previous section. This leaves a sample of 5,577 investors who hold 84,975 positions. Missing demographic information accounts for the majority of excluded positions. In fact, 70% of all positions observed in the dataset are held by an investor without occupation information, 50% each by investors without gender information or self-assessed experience, and 40% by investors without age information. The effect these variables have on the disposition effect is of primary interest, so it will still be worthwhile to fit models using this heavily reduced sample. However a model without these variables will be estimated using a much larger sample in a later section. Parameter estimates for variables that appear in both models can then be compared to see if using the reduced sample makes a significant difference. The remainder of this section and those that follow shall deal exclusively with the reduced sample unless otherwise specified.

Of the 84,975 positions in the reduced sample, 60.4% are sold within 500 days of first being purchased, with the remainder being censored at that time. As discussed in section 5, only the first sale of a position is recorded, since the holding period and reference price of subsequent sales is ambiguous. The median holding period is 155 days, and 86 days if censored observations are not included.

[8] On the date of a sale the sell price is used for this comparison, so a stock sold for a gain will always be recorded as having been trading at a gain on that date, and likewise with losses.

6 Comparison of marginal and frailty models

Following Feng and Seasholes [2005], the effect of a covariate on the DE can be measured by estimating a model for the hazard of a position being sold that includes an interaction term between the covariate and the paper status variable described in section 4. If the hazard ratio for the interaction term is greater (less) than one then the covariate makes the DE stronger (weaker). Interaction terms were initially added individually to a frailty model containing all main effects described in section 4. In this first stage, all covariates had a significant interaction with the gain indicator at the 5% level, with the exception of the SIC industry categories. Hence only the main effects for this variable were included in subsequent models.

Both a marginal and frailty model containing all main effects and gain indicator interactions were then estimated. Table 1 contains hazard ratios for interaction terms that had significant coefficient estimates in both the frailty and marginal models. Significance was tested at the 5% level using the Wald statistic. As can be seen, the hazard ratios are broadly similar in both models. Table 2 contains hazard ratios for interaction terms that had significant coefficient estimates in the frailty model but not in the marginal model. The differences between models are larger for these covariates, and in each case the hazard ratio for the marginal model is closer to one i.e. a smaller effect. However the lack of significance itself in the marginal model is the most notable difference.[9] Had a marginal model been used exclusively then the analysis would show that gender, income and diversification did not have an effect on the DE. Full regression results for the frailty interactions are provided in section 6.1, and in the appendix for the main effects. The full results for the marginal model are also provided in the appendix.

Having established that the frailty model produces materially different results

[9] In addition to those shown in tables 1 and 2, there were no interactions that were significant in the marginal model but not in the frailty model. Such a result would be unlikely in the presence of any substantial correlation at the investor level.

	HR: frailty	HR: marginal
Gain * Experience: None	1.693	1.568
Gain * Experience: Limited	1.367	1.292
Gain * Experience: Good	1.220	1.161
Gain * December	0.504	0.503
Gain * '95/'96	0.762	0.742
Gain * Cap quintile 1	1.261	1.265
Gain * Cap quintile 2	1.171	1.197
Gain * Age	0.986	0.988
Gain * Tenure 3	1.213	1.192

Table 1: Hazard ratios for interaction terms with significant estimates in both the frailty model and the marginal model.

	HR: frailty	HR: marginal
Gain * Male	1.159	1.104
Gain * Income 2	0.850	0.896
Gain * Income 3	0.902	0.931
Gain * Income 4	0.912	0.955
Gain * Diversification 2	0.924	0.983
Gain * Diversification 3	0.922	0.976
Gain * Diversification 4	0.858	0.936
Gain * Tenure 4	1.188	1.081

Table 2: Hazard ratios for interaction terms with significant estimates in the frailty model but not in the marginal model.

than the equivalent marginal model, the next step is to assess the evidence for whether the frailty model is actually a better fit for the data, and hence that its results provide a more accurate representation of what is happening in reality. The significance of the frailty component itself can be tested using a likelihood-ratio test (LRT) comparing the frailty model with the equivalent marginal version, since they differ only by the presence of the frailty component. The null hypothesis of this test is that the frailty variance is zero, which would normally be problematic since it is on the boundary of the parameter space. However, Nielsen et al. [1992] show that the usual chi-squared distribution with one degree of freedom is still valid in this case. This LRT produces a highly significant test statistic (14,918), indicating that investor-level correlation is present, and that modelling it as shared frailties is supported by the data.

	AIC	Pseudo-R^2	Concordance
Marginal	1,084,489	0.41	0.71
Frailty	1,065,433	0.66	0.79

Table 3: Statistics comparing the overall adequacy of the marginal and frailty models.

Three statistics that can be used to compare the overall fit of the models are presented in table 3. They are the AIC, a pseudo-R^2 statistic due to Xu and O'Quigley [1999] and the concordance. The likelihood used in calculating the AIC and pseudo-R^2 for the frailty model has had the frailty component integrated out, so that it is comparable with the likelihood from the marginal model. The concordance is the proportion of all pairs of observations for which the model assigns greater hazard to the one that experiences an event first. A concordance over 70% is good for any survival model, so it is reassuring that both models are able to achieve this level. The frailty model performs better in each of these three measures, providing strong evidence that the inclusion of a frailty component produces a better fit to the data. The adjustment to standard errors in the marginal model does not affect these three statistics, hence they are the same as what would be obtained if a naive Cox model with no adjustment had been estimated.

Another important comparison to make is the degree to which the proportional hazards (PH) assumption holds in each model. If the effect of a variable appears significantly non-proportional, then the hazard ratio for that variable is not a good summary of its effect and care needs to be taken when interpreting it. The validity of the PH assumption can be checked for a particular variable by calculating residuals introduced by Schoenfeld [1982]. Grambsch and Therneau [1994] propose a graphical method for checking the PH assumption where the scaled Schoenfeld residuals (SSR) for a covariate are plotted against the survival times, and a smoothed line fitted to these points. The implementation in the 'survival' package also plots an approximate 95% confidence interval around the curve. A non-zero slope in this curve indicates a deviation from PH, with a positive slope implying the hazard ratio for this covariate increases with survival time and a negative slope implying that it decreases.

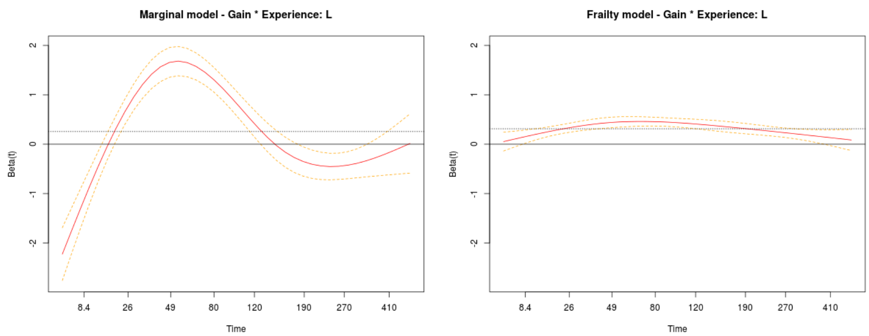

Figure 1: Plot of a smooth of the scaled Schoenfeld residuals for the interaction term between the paper gain indicator and the 'Limited' group for self-assessed experience, in the frailty and marginal models. Each curve is plotted with an approximate 95% confidence interval. A horizontal dashed line is plotted at the coefficient estimate for the variable in the respective model.

Figures 1 and 2 show SSR plots for the interactions with the gain indicator for the 'Limited' group of the experience variable, and age respectively in the frailty model and marginal models. A horizontal dashed line is plotted in each figure (and all SSR figures below) at the parameter estimate for the variable. Clearly, there is a large departure from proportional hazards in both plots for the marginal model. The interaction for age in the frailty model did not deviate much anyway, but for the experience group the deviation that is there is small compared to what is observed in the marginal model.

Overall adherence to the PH assumption can be assessed by computing the correlation between the SSR and the survival times for each of the interaction terms, as described in Grambsch and Therneau [1994]. The median of the absolute values of these correlations for the frailty model is 0.0039, compared to 0.0193 for the marginal model. The authors also derive a test statistic for a non-zero slope that

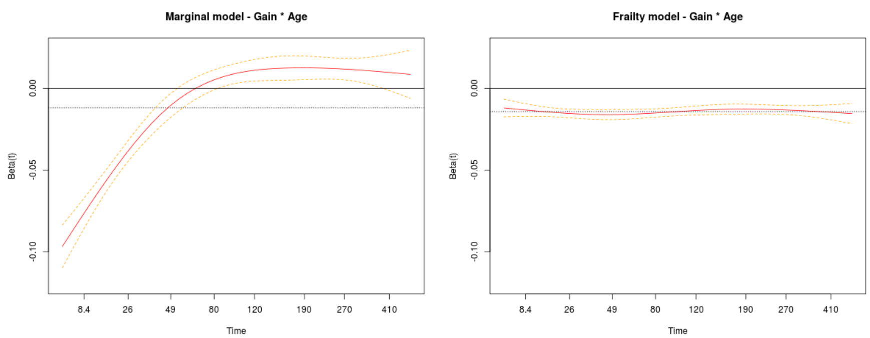

Figure 2: Plot of a smooth of the scaled Schoenfeld residuals for the interaction term between the paper gain indicator and age, in the final frailty model and an equivalent marginal model. Each curve is plotted with an approximate 95% confidence interval. A horizontal dashed line is plotted at the coefficient estimate for the variable in the respective model.

is asymptotically χ^2 under the null of no correlation. The null is rejected at the 1% level for 19/22 of the interaction terms in the marginal model, with test statistics that are at least one order of magnitude larger than their counterparts in the frailty model. In the frailty model, 8/22 have non-zero correlation significant at the 1% level. These results are displayed in full in the appendix. Since the sample size is large, highly significant test statistics are more likely. But interpretation of a covariate's effect on the DE may not change much as a result. Reference to the smoothed SSR plots is therefore an important step when estimated coefficients are being interpreted, as they show the nature of the deviation from PH.

These results highlight two potential advantages of using a frailty model for measuring the effect of covariates on the DE in a dataset of this kind. Firstly, it has been possible to significantly estimate effects which would otherwise have been obscured in a marginal model. Secondly, doing so has greatly reduced the deviation from proportional hazards in the interaction terms, meaning the hazard ratios for these terms are good summaries of the effects and can be reported with confidence.

6.1 Frailty model results

	HR	SE	P-value
Gain * Professional occupation	0.99	0.02	0.57
Gain * Male	1.16	0.04	0.00
Gain * Experience: N	1.69	0.07	0.00
Gain * Experience: L	1.37	0.03	0.00
Gain * Experience: G	1.22	0.02	0.00
Gain * Income 2	0.85	0.03	0.00
Gain * Income 3	0.90	0.03	0.00
Gain * Income 4	0.91	0.03	0.00
Gain * December	0.50	0.03	0.00
Gain * 95/96	0.76	0.02	0.00
Gain * Cap quintile 1	1.26	0.07	0.00
Gain * Cap quintile 2	1.17	0.04	0.00
Gain * Cap quintile 3	1.06	0.03	0.10
Gain * Cap quintile 4	0.98	0.03	0.43
Gain * Age	0.99	0.00	0.00
Gain * Diversification 2	0.92	0.02	0.00
Gain * Diversification 3	0.92	0.03	0.01
Gain * Diversification 4	0.86	0.03	0.00
Gain * Log sales made	0.99	0.01	0.13
Gain * Tenure 2	1.03	0.03	0.21
Gain * Tenure 3	1.21	0.03	0.00
Gain * Tenure 4	1.19	0.03	0.00

Table 4: Hazard ratios, standard errors and p-values for the interaction terms in the frailty model with all interaction terms included.

This section will discuss the results for each interacted variable in the frailty model, as shown in table 4. Hazard ratios and Wald test p-values for the main effects are provided in the appendix. As well as the hazard ratio and its statistical significance, the presence of any non-proportionality in the effect of an interaction term will also be considered, as this may change the interpretation.

In this full model, the interactions for the professional occupation indicator and the log of number of sales made to date did not have significant estimates. In the case of the sales made variable, the estimate retains its significance in a model estimated without the self-assessed experience variable. In this model, the hazard ratio is 0.977. The IQR for log of number of sales made is about 2.5, so a of this magnitude corresponds to a hazard ratio of 0.943 ($0.977^{2.5}$) and a reduction in hazard of roughly 6%. Although not strong, there is some association between the two variables. Regressing log sales on experience at the interval level[10] produces an R^2 of 0.04. Since the effect of log sales is small anyway in the model without experience, the addition of experience appears to be enough to wipe it out. The professional occupation indicator still does not have a significant estimate in a model without experience, and it is not strongly correlated with any other variable. When tested individually, the hazard ratio for the indicator was 1.09, meaning the DE was stronger for investors who worked in a professional occupation. This is contrary to what was expected based on past research. Due to its insignificance in the full model though, the conclusion made here must be that it does not have an effect when other factors are controlled for.

Moving on to variables that do have a significant effect on the DE, the hazard ratio for the gender interaction is 1.16, so on average the DE is 16% stronger for men. However this is a case where the SSR plot reveals large fluctuations of the smoothed curve for $\beta(t)$ around the estimate (indicated by the red curve and black dotted line respectively), as shown in figure 3. Whilst the effect is not constant over time, it is consistently positive. Hence the conclusion of a difference between genders in this regard still seems valid, although the magnitude is not constant as holding period increases.

The experience variable has both highly significant main effects and interaction terms, with more experienced investors selling positions at a greater rate in general and exhibiting less of a DE. Compared to the 'extensive' group, the 'good', 'limited' and 'none' groups have a DE that is 22%, 37% and 69% stronger respectively. This is the largest effect amongst all of the categorical variables. The SSR

[10]Intervals are blocks of time during the holding period of a position where none of the time-varying covariates change.

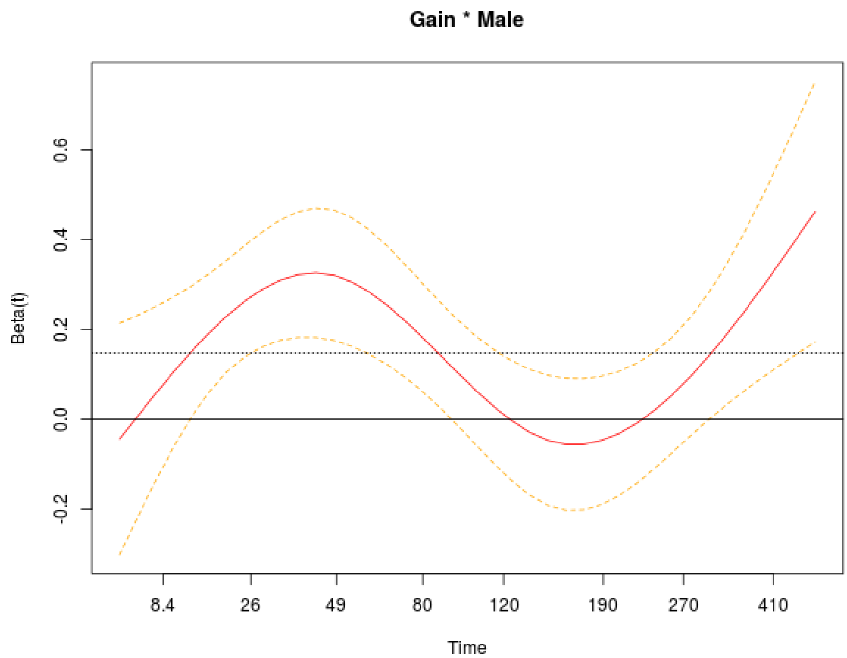

Figure 3: Plot of a smooth of the scaled Schoenfeld residuals for the interaction term between the paper gain indicator and the gender indicator (equalling one if the investor is male) in the frailty model. An approximate 95% confidence interval for the curve is also plotted. A horizontal dashed line is plotted at the coefficient estimate.

plots show a peak in the effect after around 80 days for the 'limited' and 'good' categories, as shown in a plot of the former in figure 4 (the shape for the 'good' category is similar, but less severe). This means that the difference between these categories and the 'extensive' category takes some time to fully materialize, and also diminishes after the holding period has reached a certain length.

For age, an increase of one year corresponds to a 3.4% reduction in DE, and the IQR of 16 years corresponds to a 20% reduction. The SSR plot does not show any sign of non-proportionality, hence the hazard ratio estimate is a good summary of the effect. The results for the income variable show that income group 1 (lowest income) has the largest DE, as expected. But the largest reduction in DE is

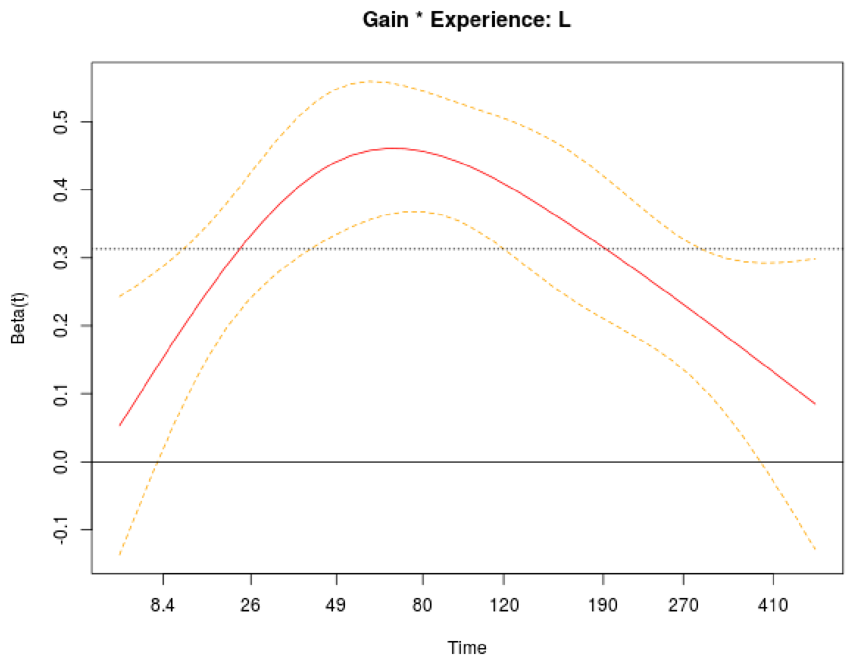

Figure 4: Plot of a smooth of the scaled Schoenfeld residuals for the interaction term between the paper gain indicator and the 'limited' experience category in the frailty model. An approximate 95% confidence interval for the curve is also plotted. A horizontal dashed line is plotted at the coefficient estimate.

exhibited by income group 2 (15% reduction), rather than the two higher income groups (9% reductions for both). For group 3, the SSR plot shows a stronger effect for the first 50 days of a holding period, during which it is closer to that of group 2, before decreasing and remaining stable for the rest of the follow-up time.

Investors with a greater level of portfolio diversification had a weaker DE on average, with those holding at least 12 different stocks having a 14% reduction in DE compared to those holding 1 or 2. Holding between 3 and 12 stocks lead to an 8% reduction, with no difference between those holding more or less than 7 within this range. Only the SSR plot for the third group shows cause for concern, with the effect falling to essentially zero between roughly 40 and 80 days, as shown in figure 5.

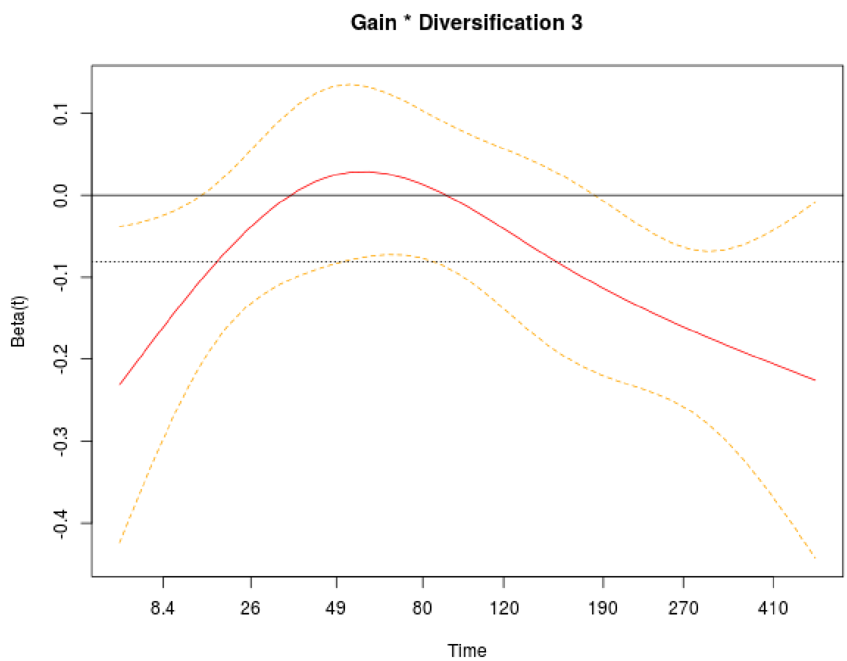

Figure 5: Plot of a smooth of the scaled Schoenfeld residuals for the interaction term between the paper gain indicator and the third diversification group (holding between 7 and 11 stocks) in the frailty model. An approximate 95% confidence interval for the curve is also plotted. A horizontal dashed line is plotted at the coefficient estimate.

Only tenure groups 3 and 4 (first account opened between 4 and 7 years prior to start of data period, and first account opened at least 8 years prior respectively) have significantly different effects on the DE than group 1 (account opened between one year prior and one year following the start of the data period). Both exhibit a DE that is 20% larger than group 1, meaning that investors who have held an account at this brokerage for at least 4 years have a stronger DE than those who have held one for less. This result will be discussed further in section 8.

As expected, the DE is reduced in December months, in fact by 50% as shown by the hazard ratio. As shown in figure 6, the SSR plot shows the effect increasing from -0.2 (hazard ratio = 0.82) to almost -1.2 (hazard ratio = 0.3) by the end of the follow-up time. This means that the reduction in DE in December is much larger for positions that have already been held for a longer period of time. This is an important caveat to the finding of a significant December effect that would not have been revealed without inspecting the SSR plot.

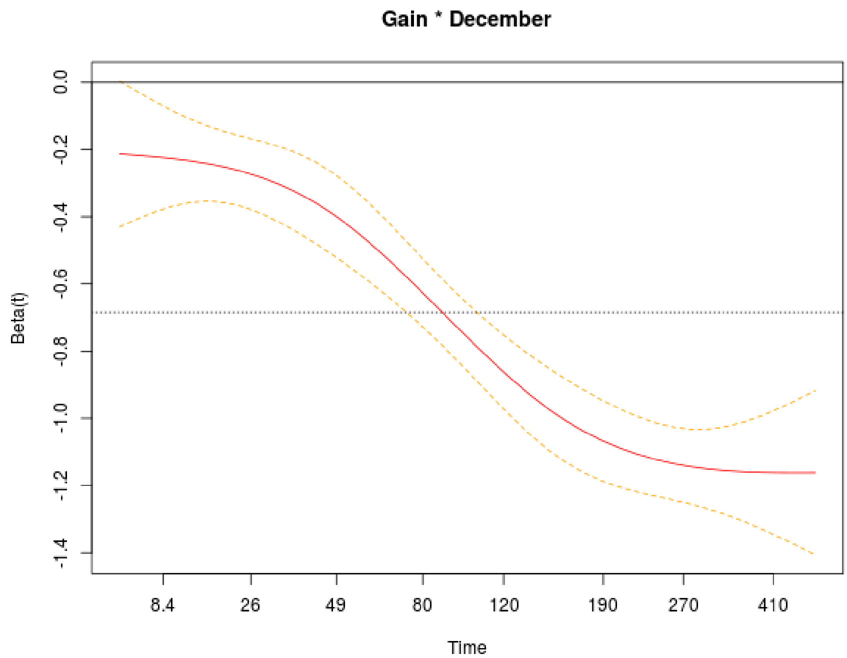

Figure 6: Plot of a smooth of the scaled Schoenfeld residuals for the interaction term between the paper gain indicator and the December indicator in the frailty model. An approximate 95% confidence interval for the curve is also plotted. A horizontal dashed line is plotted at the coefficient estimate.

Confirming the exploratory analysis, in this model the DE is significantly lower in years 1995 and 1996, with the hazard ratio indicating an average reduction of 24%. The SSR plot shows that the effect is stronger at the start of the follow-up time and weaker towards the end, but it stays close enough to the parameter estimate such that the hazard ratio is still a reasonable summary of the effect.

Stocks in the bottom two quintiles of capitalization size have an increased DE relative to those in the largest cap-size quintile. There is an increase of 26% for the first quintile and 17% for the second. There is not a significant difference between the largest quintile and quintiles 3 and 4. Figures 7 and 8 show SSR plots for these interaction terms. Importantly, the effect is not significantly above zero until roughly 100 days into the follow-up time in both cases, and continues increasing after this point. Since the interaction terms for quintiles 3 and 4 were not significant, this means that cap-size does not have an effect on the DE until roughly 100 days into a holding period.

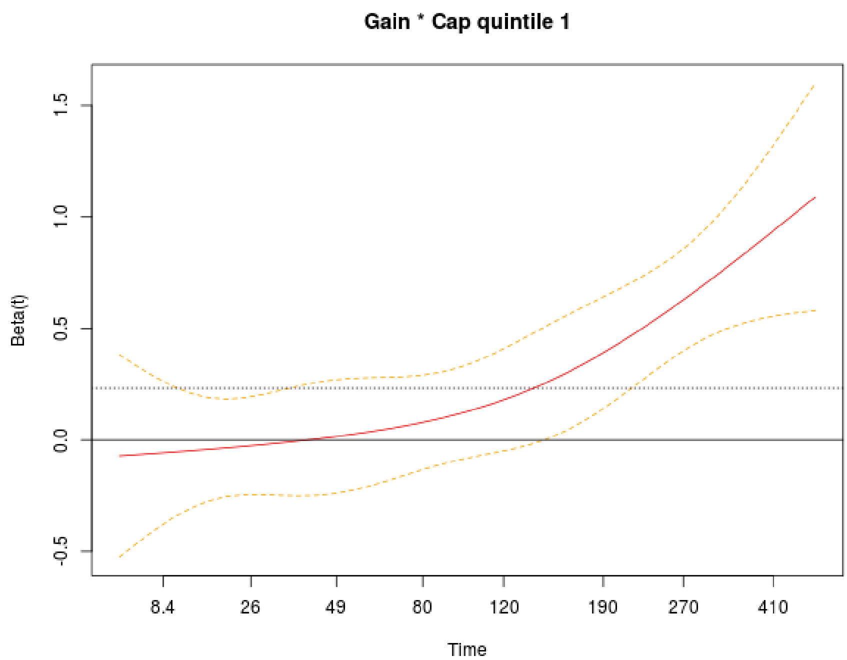

Figure 7: Plot of a smooth of the scaled Schoenfeld residuals for the interaction term between the paper gain indicator and the first capitalization size quintile, containing stocks with the smallest cap-sizes.

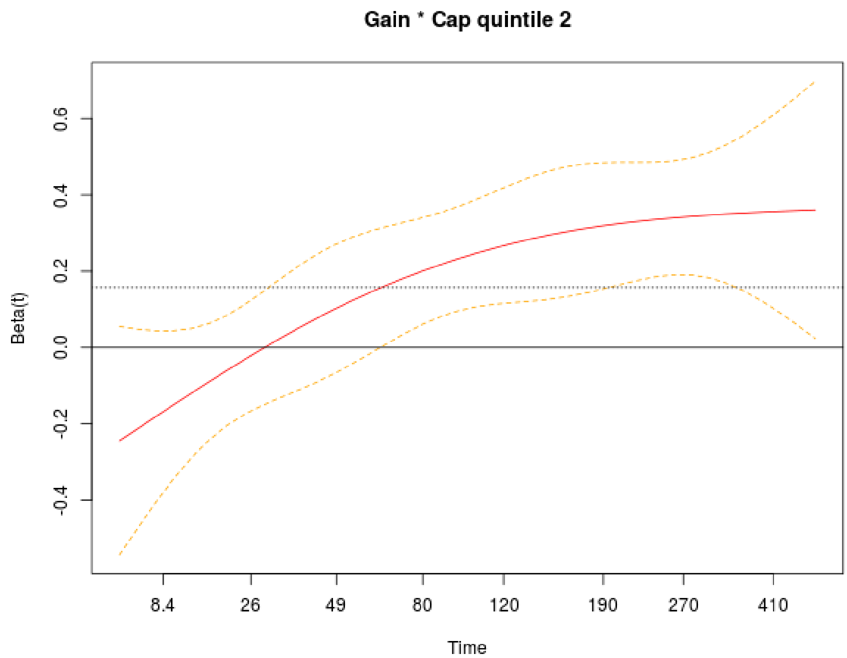

Figure 8: Plot of a smooth of the scaled Schoenfeld residuals for the interaction term between the paper gain indicator and the second capitalization size quintile.

7 Robustness checks and supplementary models

7.1 Model without demographic information

One thing that distinguishes the LDB dataset from others used in the literature is the large range of demographic information that is available for some investors. The improvement in the model due to the inclusion of demographic information is therefore of interest. To answer this, the frailty model from section 6.1 was re-estimated without any of the demographic variables or their interactions. This includes age, gender, the professional occupation indicator, self-assessed experience, and income group. The results for the variables common to both models are very similar, in sign, magnitude and significance. The only substantive difference is that the sales made interaction is significant in the model without the demographic variables. As was mentioned in section 6.1, the log sales made variable was significant if the experience variable was excluded from the model, hence the result here is not surprising.

The two models can also be compared in their overall adequacy. As would be expected, the model with demographic variables is a better fit, judged by a few different measures, but the difference is small. The model with demographic variables has concordance of 0.791 compared to 0.790 for the model without, AIC of 1,056,616 compared to 1,057,150 and pseudo-R^2 of 0.655 compared to 0.652. The effect sizes for the demographic variables are fairly small, with the exception of self-assessed experience, so the model would not be expected to suffer greatly in their absence. But when the model was re-estimated without these variables, new frailty values were also computed, and since all the demographic variables are static over time, it is likely that the frailty component was able to do a good job of adjusting for the additional heterogeneity introduced by the their omission. This provides some evidence that the frailty model is also able to adjust for sources of heterogeneity that are truly unobserved.

7.2 Model without demographic variables and larger sample

In order to include demographic variables in the model, the sample had to be restricted to investors for whom full demographic information was available. This meant the sample used to fit the final model in section 6.1 contained 5,577 investors and 84,975 positions. This is compared to the 55,000 investors with 618,000 positions recorded in the full dataset. It is possible that the small sample with demographic variables is biased in some way that would result in different parameter estimates for the non-demographic variables compared to what would be obtained with the full sample. To check for this, the model from the previous section was re-estimated but with no demographic variables and the largest possible sample. Due to memory constraints, this contained 364,664 randomly sampled positions.[11]

Sample size	Small	Large	Small	Large
	HR	HR	P-value	P-value
Gain * December	0.504	0.540	0.000	0.000
Gain * 95/96	0.762	0.812	0.000	0.000
Gain * Cap quintile 1	1.263	1.433	0.001	0.000
Gain * Cap quintile 2	1.170	1.207	0.000	0.000
Gain * Cap quintile 3	1.055	1.106	0.104	0.000
Gain * Cap quintile 4	0.980	1.078	0.419	0.000
Gain * Diversification 2	0.922	0.904	0.001	0.000
Gain * Diversification 3	0.919	0.908	0.003	0.000
Gain * Diversification 4	0.858	0.790	0.000	0.000
Gain * Log sales made	0.989	0.941	0.143	0.000
Gain * Tenure 2	1.035	1.033	0.178	0.007
Gain * Tenure 3	1.218	1.061	0.000	0.000
Gain * Tenure 4	1.190	1.099	0.000	0.000

Table 5: Hazard ratios for interaction terms in a frailty model that does not include any demographic information, and estimated using a much larger sample (364,000 positions compared to 85,000). Compared with hazard ratios for the same interactions as estimated in the full frailty model from section 6.1, which used the smaller sample.

Table 5 shows the hazard ratio and estimate significance for each interaction term that is common to both the model from section 6.1, which also contained demographic variables, and the model estimated using the much larger sample. The hazard ratios for the December and '95/'96 indicators are very similar and both highly significant. For cap-size quintiles the pattern in hazard ratios is similar with positions in stocks from smaller cap-size quintiles having a stronger DE. However the differences between quintiles 3, 4 and 5 were significant in the

[11]Although the full dataset of positions itself is only 1.3GB in size, the procedure for fitting a frailty model is very memory intensive. A machine with 16GB of memory was able to fit a model using half of the full dataset, hence half of all positions were randomly sampled. Fitting a model using this sample took 22 hours.

larger sample size model, but were not in the small sample size model.

There is again close agreement for diversification, with groups 2 and 3 having a small decrease in DE relative to group 1, and group 4 having a larger relative decrease. As expected from the supplementary models estimated already, log of sales made is highly significant in the larger sample model, with a hazard ratio of 0.941. For an increase of 10 sales made this translates into a reduction in DE by 13%, and a reduction by 17% for an increase of 20 sales made. For investor tenure, the large sample model again finds that it is investors who have had an account for the least amount of time (tenure group 1) that have the lowest DE, with tenure groups 2, 3 and 4 having a DE that is 3%, 6% and 10% than group 1 respectively. In the small sample model the increase for groups 3 and 4 was much larger, around 20% in both cases. As discussed previously, the increase in DE for investors who have held accounts for longer is the opposite of what was expected, and the reason for it remains unknown. It could be that tenure is capturing something else, such as an investor's willingness to 'shop around' for the best terms when choosing which brokerage to trade with. Shorter tenure may suggest a greater willingness to switch for a better deal, which may go along with a more sophisticated approach to investing, and hence a lower DE.

8 Discussion

Section 6 showed that the addition of a frailty component significantly improved the Cox model for position lifetimes, in terms of both goodness-of-fit and adherence to the PH assumption. By isolating the effect of covariates and their interactions from the unobserved frailty unique to each investor, many more of these effects could be significantly estimated in the frailty model. This is of particular importance in a sample like this where the distribution of positions amongst investors is extremely imbalanced.

That the frailty model is able to control for unobserved heterogeneity is supported by the results of section 7.1, where a model was fitted using the same sample as the full interaction model, but without any demographic variables. The parameter estimates for variables common to both models did not change much, which would be a concern when variables known to be important are omitted. The overall performance of the model did also not greatly suffer, with concordance and pseudo-R^2 values close to the full model. This suggests the frailty terms were able to 'absorb' some of the static differences explained by the demographic variables when the model was re-estimated without them.

Graphically checking the PH assumption proved to be an important step in the modelling process. Deviations from PH substantially changed the interpretation of some effects compared to what would have been concluded if only the hazard ratios were considered. For example, the reduction in DE in December was shown to be much stronger for positions that had already been held for a long period of time. This adds importance nuance to the evidence on this topic.

8.1 Model results

This section will summarise the results of the analysis, making comparisons to those in Feng and Seasholes [2005] (F&S) and [Dhar and Zhu, 2006] (D&Z) when possible.

8.1.1 Demographic covariates

The DE for men was 16% stronger on average in this sample. This is in contrast to F&S who find a DE that is 30% stronger for women, with a much more even gender balance: 51% men in their dataset from China compared to 80% in the LDB data. This suggests that gender differences in the prevalence of the DE vary between countries, but more research on this specific topic would be needed to understand the nature and causes of this variation. Despite a significantly weaker DE being found amongst those working in a professional occupation by D&Z,

no such difference was found here. This lack of significance persists when the self-assessed experience variable is removed from the model.

For age, an increase of one year corresponds to a decrease in DE by 1.5%, or for the IQR of 16 years a decrease by 20%. D&Z also find that older investors suffer less from the DE, although comparing the magnitude of the effects is difficult due to the difference in methodology. Rather than greater age being associated with increased investing sophistication, F&S hypothesise, and confirm in their results, that younger investors in China will be more sophisticated due to older investors having grown up in a radically different economic system.

Those with incomes in the lowest of the sample quartiles have the strongest DE, but there is not a monotonic decrease for higher quartiles. D&Z split investors into three income groups, and find a 10% reduction in DE for the highest compared to the lowest. Their low and high groups correspond closely to the lowest and highest quartiles used here, for which a 9% difference was found. The results in this analysis show that there are a significant number of investors with incomes in the top two quartiles who suffer from a stronger DE and therefore bring down the group average. Assuming more sophisticated investors will exhibit a lower DE, these results show that income is not a good measure of sophistication when other factors are controlled for.

8.1.2 Other investor-level covaraites

Self-assessed experience proved to have a lot of explanatory power when it came to differences in the disposition effect. Investors with the lowest level of experience had a DE that was 69% stronger than that of the group with the highest level. It is interesting that investors' assessment of their own ability matched up so well with the extent to which they committed what is generally considered to be an investment mistake i.e. selling winners too soon and holding losers too long. This raises the question of why investors who consider themselves to be inexperienced trade at all.

It may be that they hope to learn by trading and improve their performance as they gain in experience. The results here provide mixed evidence for this. When self-assessed experience is not included in the model, an increase in the number of trades an investor has made reduces their DE, with an increase of 10 trades corresponding to a reduction in DE by 13%. Importantly, only trades made during the data period can be counted, and for some investors this will represent only a small portion of their investing career. However, when self-assessed experience is included in the model, the number of sales an investor has made no longer has a significant effect. This suggests that the number of sales made is actually capturing differences that existed before the start of the data period, and not learning that happens as a result of the trading that's observed.

One explanation for this is that the investors who rate themselves as having high experience hold and sell more positions during the data period. In fact the group with the highest self-assessed experience contains 14% of investors, but account for 22% of all positions. The number of sales made variable will increase faster for investors in this group, and throughout the data period larger values for number of sales made will typically indicate an investor has a higher level of self-assessed experience. F&S and D&Z also found that trading more reduced an investors DE, but did not have a similar measure of self-assessed experience in their models. This provides some new evidence on the question of whether investors learn from trading, and shows that the number of trades an investor has made is perhaps not a good measure of gained experience by itself.

Investors with a greater level of initial diversification had a weaker DE. It was reduced by 14% for those holding at least 12 stocks initially, compared to those holding less than 3. Investors initially holding 3-11 stocks had a DE 8% lower than those holding less than three. F&S found that holding at least two stocks corresponded to a 16% reduction in DE.

Somewhat surprisingly, those who had held an account for the shortest time had the weakest DE. The increases in DE for groups who had held an account for longer were smaller, but still present, in the model estimated using the much larger sample. This is contrary to what was expected, with one explanation being that account tenure is not measuring investor experience or sophistication in this dataset. Tenure is positively correlated with age, but not strongly so (Pearson's correlation coefficient of 0.15), and the mean value of tenure is only 4 years. This suggests there is not a strong relationship between account tenure and the amount of time an investor has been investing in total. If there was, we would expect more large values of tenure, and large values would only be possible for older investors, hence stronger correlation with age.

This result is also the opposite of what would be expected if investors with lower investing ability were dropping out as a result of their poor performance. Assuming that better investors will suffer less from the DE, and that poor performance makes an investor more likely to close their account[12], then the investors who have held accounts the longest should have lower DE as a group. The decision to cease trading or even close an account is likely more complicated than whether an investor's trading performance has been good or not, which itself is affected by more than just the extent to which they have exhibited the DE. Further research would be needed to untangle the different factors in this issue.

8.1.3 Stock-level and calendar covariates

The DE was greater for positions in stocks with a smaller cap-size. In the model with demographic variables there was not a significant difference between cap-size quintiles 3-5, but a difference between these groups was detected in the model estimated with a much larger sample size in section 7.2. In this latter model, the increases in DE relative to the fifth quintile (stocks with the largest cap-sizes) were 43%, 21%, 11% and 8% respectively for quintiles 1-4. It was important in this case to check the SSR plots, as they showed that the effect for all quintiles does not really take effect until 100 days into the follow up time for a position.

[12] Or at least stop actively trading common stocks, in which case they would not appear in the sample used here, even if they do not close their account at the brokerage entirely.

This is evidence that investors consider different characteristics of stocks they have positions in at different times when deciding whether to sell the position or not.

Smaller cap-size stocks tend to be more volatile and Kumar [2009] hypothesises that more volatile stocks are harder for investors to value, which makes them more likely to exhibit a DE when trading small cap stocks. So one possibility is that the volatility of a stock only begins to affect an investor's decision making after they have experienced it first-hand and seen the fluctuations in price as they hold it. Unlike cap-size, the industry group of the stock being held did not have an effect on the DE. There were differences between groups in terms of their hazard of being sold, which merited this variables inclusion as a main effect. But the interactions with the gain indicator were not significant, even when tested in a model without interactions for any other variables.

As expected, the DE was much weaker during December months, being reduced by half compared to other times of the year. The SSR plot showed that the effect increased dramatically over time, from -0.2 (hazard ratio = 0.82) to almost -1.2 (hazard ratio = 0.3) by the end of the follow-up time. This means that, when wanting to sell a losing position in December for tax purposes, investors sell 'old' losses at a greater rate than losing positions they have not held for very long. One explanation is that investors are more optimistic about the possibility of a price reversal in positions that are still relatively new. This result adds nuance to our understanding of the December effect on the DE, and provides evidence that investors do consider how long they have held a stock when deciding whether to sell it or not. The result also raises questions that cannot be answered by this model. It is not possible to tell whether investors are actively choosing to sell older losses since this would require knowing if they have a younger loss available to sell as well. It may also be the case that investors prefer to sell positions that are further from the purchase price when they have a choice, and older positions are more likely to have fallen further in price.

In this dataset the DE is significantly lower in years 1995 and 1996, with the hazard ratio indicating an average reduction of 24%. The SSR plot shows that the effect is stronger at the start of the follow-up time i.e. for stocks that have only recently been purchased, and weaker towards the end, but it stays close enough to the parameter estimate such that the hazard ratio is still a reasonable summary of the effect. These results establish that the fall in aggregate DE during these years, as observed in Kumar [2009], cannot be entirely explained by other factors, such as the characteristics of investors trading in these final two years of the data period, since they are controlled for in the model. The decrease in DE is therefore likely due to changes in market conditions during the data period.

Appendix

	Frailty			Marginal		
	Correlation	χ^2	P-value	Correlation	χ^2	P-value
Gain * Professional occupation	-0.002	0.230	0.631	-0.027	392.837	0.000
Gain * Male	-0.002	0.137	0.711	-0.008	48.200	0.000
Gain * Experience: N	-0.004	0.990	0.320	-0.025	539.597	0.000
Gain * Experience: L	-0.002	0.255	0.613	-0.001	0.913	0.339
Gain * Experience: G	-0.001	0.069	0.793	-0.002	2.148	0.143
Gain * Income 2	-0.003	0.672	0.413	-0.004	7.140	0.008
Gain * Income 3	0.004	0.735	0.391	0.021	219.909	0.000
Gain * Income 4	-0.001	0.079	0.778	0.017	186.775	0.000
Gain * December	-0.048	119.957	0.000	-0.031	133.792	0.000
Gain * 95/96	0.014	10.207	0.001	0.035	321.657	0.000
Gain * Cap quintile 1	0.019	19.016	0.000	0.037	184.752	0.000
Gain * Cap quintile 2	0.017	15.200	0.000	0.035	273.665	0.000
Gain * Cap quintile 3	0.019	19.737	0.000	0.035	388.410	0.000
Gain * Cap quintile 4	0.024	30.948	0.000	0.036	247.704	0.000
Gain * Age	0.002	0.129	0.719	0.022	193.935	0.000
Gain * Diversification 2	0.003	0.374	0.541	0.002	2.402	0.121
Gain * Diversification 3	-0.004	0.983	0.322	0.007	22.437	0.000
Gain * Diversification 4	-0.005	1.154	0.283	0.004	10.244	0.001
Gain * Log sales made	0.013	8.829	0.003	-0.038	1145.005	0.000
Gain * Tenure 2	0.020	21.963	0.000	0.018	183.899	0.000
Gain * Tenure 3	0.000	0.001	0.972	-0.009	42.030	0.000
Gain * Tenure 4	0.004	0.724	0.395	-0.014	121.455	0.000

Table 6: Correlation coefficient between the scaled Schoenfeld residuals and survival times for each interaction term in the marginal and frailty models. Reported with the χ^2 test statistics for whether this correlation is non-zero and associated p-values.

	HR	SE	P-value
Neutral	1.14	0.02	0.00
Gain	4.21	0.07	0.00
Professional occupation	1.06	0.03	0.08
Male	1.04	0.06	0.49
Experience: N	0.57	0.11	0.00
Experience: L	0.62	0.05	0.00
Experience: G	0.75	0.05	0.00
Income 2	1.04	0.04	0.41
Income 3	1.12	0.04	0.01
Income 4	1.03	0.06	0.64
December	1.62	0.02	0.00
95/96	1.18	0.02	0.00
Cap quintile 1	0.78	0.05	0.00
Cap quintile 2	0.86	0.03	0.00
Cap quintile 3	1.01	0.02	0.81
Cap quintile 4	1.11	0.02	0.00
Age	1.00	0.00	0.00
Diversification 2	0.98	0.04	0.66
Diversification 3	0.85	0.05	0.00
Diversification 4	0.81	0.06	0.00
Log sales made	1.17	0.01	0.00
Tenure 2	0.99	0.04	0.82
Tenure 3	0.79	0.04	0.00
Tenure 4	1.00	0.06	0.99
SIC: AFF	1.09	0.23	0.72
SIC: CON	1.00	0.06	0.94
SIC: FIRE	0.78	0.02	0.00
SIC: MIN	0.81	0.03	0.00
SIC: PUBA	1.19	0.13	0.18
SIC: RETL	0.92	0.02	0.00
SIC: SERV	1.03	0.01	0.02
SIC: TRAN	0.69	0.02	0.00
SIC: WHOL	1.10	0.03	0.00

Table 7: Hazard ratios, standard errors and p-values for the main effects in the frailty model with all interaction terms included.

	HR	SE	P-value
Neutral	1.087	0.026	0.001
Gain	3.603	0.194	0.000
Professional occupation	1.093	0.041	0.028
Male	1.024	0.092	0.795
Experience: N	0.684	0.135	0.005
Experience: L	0.809	0.063	0.001
Experience: G	0.892	0.052	0.030
Income 2	1.009	0.054	0.863
Income 3	0.998	0.055	0.974
Income 4	0.915	0.066	0.183
December	1.594	0.034	0.000
95/96	0.854	0.031	0.000
Cap quintile 1	0.868	0.056	0.011
Cap quintile 2	0.925	0.040	0.049
Cap quintile 3	1.037	0.033	0.264
Cap quintile 4	1.142	0.024	0.000
Age	1.004	0.002	0.005
Diversification 2	0.866	0.047	0.002
Diversification 3	0.671	0.062	0.000
Diversification 4	0.556	0.063	0.000
Log sales made	1.640	0.016	0.000
Tenure 2	1.105	0.052	0.057
Tenure 3	0.907	0.051	0.056
Tenure 4	1.189	0.068	0.011
SIC: AFF	1.405	0.228	0.136
SIC: CON	0.992	0.058	0.893
SIC: FIRE	0.775	0.022	0.000
SIC: MIN	0.823	0.033	0.000
SIC: PUBA	1.099	0.131	0.472
SIC: RETL	0.949	0.020	0.010
SIC: SERV	1.042	0.019	0.027
SIC: TRAN	0.696	0.023	0.000
SIC: WHOL	1.127	0.030	0.000

Table 8: Hazard ratios, robust standard errors and Wald test p-values for the main effects in the marginal model. Standard errors are robust to correlation at the investor level.

	HR	SE	P-value
Gain * Professional occupation	0.974	0.053	0.623
Gain * Male	1.104	0.122	0.417
Gain * Experience: N	1.568	0.223	0.044
Gain * Experience: L	1.292	0.080	0.001
Gain * Experience: G	1.161	0.069	0.032
Gain * Income 2	0.896	0.071	0.123
Gain * Income 3	0.931	0.071	0.313
Gain * Income 4	0.955	0.091	0.614
Gain * December	0.503	0.042	0.000
Gain * 95/96	0.742	0.037	0.000
Gain * Cap quintile 1	1.265	0.077	0.002
Gain * Cap quintile 2	1.197	0.056	0.001
Gain * Cap quintile 3	1.090	0.045	0.055
Gain * Cap quintile 4	1.018	0.032	0.574
Gain * Age	0.988	0.002	0.000
Gain * Diversification 2	0.983	0.061	0.775
Gain * Diversification 3	0.976	0.074	0.744
Gain * Diversification 4	0.936	0.084	0.432
Gain * Log sales made	0.979	0.022	0.331
Gain * Tenure 2	1.035	0.070	0.623
Gain * Tenure 3	1.192	0.070	0.012
Gain * Tenure 4	1.081	0.090	0.386

Table 9: Hazard ratios, robust standard errors and Wald test p-values for the interactions in the marginal model. Standard errors are robust to correlation at the investor level.

References

Brad M Barber and Terrance Odean. Trading is hazardous to your wealth: The common stock investment performance of individual investors. *The Journal of Finance*, 55(2):773–806, 2000.

Brad M Barber and Terrance Odean. Boys will be boys: Gender, overconfidence, and common stock investment. *The Quarterly Journal of Economics*, 116(1): 261–292, 2001.

Philip Brown, Nick Chappel, Ray da Silva Rosa, and Terry Walter. The reach of the disposition effect: Large sample evidence across investor classes. *International Review of Finance*, 6(1-2):43–78, 2006.

David R Cox et al. Regression models and life tables. *JR stat soc B*, 34(2): 187–220, 1972.

Ravi Dhar and Ning Zhu. Up close and personal: Investor sophistication and the disposition effect. *Management Science*, 52(5):726–740, 2006.

Bradley Efron. The efficiency of cox's likelihood function for censored data. *Journal of the American statistical Association*, 72(359):557–565, 1977.

Lei Feng and Mark S Seasholes. Do investor sophistication and trading experience eliminate behavioral biases in financial markets? *Review of Finance*, 9(3):305–351, 2005.

William N Goetzmann and Alok Kumar. Equity portfolio diversification*. *Review of Finance*, 12(3):433–463, 2008.

Patricia M Grambsch and Terry M Therneau. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*, 81(3):515–526, 1994.

Mark Grinblatt and Matti Keloharju. What makes investors trade? *The Journal of Finance*, 56(2):589–616, 2001.

John P Klein. Semiparametric estimation of random effects using the cox model based on the em algorithm. *Biometrics*, pages 795–806, 1992.

Alok Kumar. Hard-to-value stocks, behavioral biases, and informed trading. *Journal of Financial and Quantitative Analysis*, 44(06):1375–1401, 2009.

Danyu Y Lin and Lee-Jen Wei. The robust inference for the cox proportional hazards model. *Journal of the American statistical Association*, 84(408):1074–1078, 1989.

Stuart R Lipsitz, Michael Parzen, et al. A jackknife estimator of variance for cox regression for correlated survival data. *Biometrics*, 52(1):291, 1996.

Gert G Nielsen, Richard D Gill, Per Kragh Andersen, and Thorkild IA Sørensen. A counting process approach to maximum likelihood estimation in frailty models. *Scandinavian journal of Statistics*, pages 25–43, 1992.

Terrance Odean. Are investors reluctant to realize their losses? *The Journal of finance*, 53(5):1775–1798, 1998.

David Schoenfeld. Partial residuals for the proportional hazards regression model. *Biometrika*, 69(1):239–241, 1982.

Terry M Therneau. *A Package for Survival Analysis in S*, 2014. URL `http://CRAN.R-project.org/package=survival`. R package version 2.37-7.

Terry M Therneau and Patricia M Grambsch. *Modeling survival data: extending the Cox model*. Springer Science & Business Media, 2000.

Andreas Wienke. *Frailty models in survival analysis*. CRC Press, 2010.

Ronghui Xu and John O'Quigley. An r2 type measure of dependence for proportional hazards models. *Journal of Nonparametric Statistics*, 12(1):83–107, 1999.