Survey Methodology, June 2007
Vol. 33, No. 1, pp. 45-50
Statistics Canada, Catalogue No. 12-001

Ex post weighting of price data to estimate depreciation rates

Marc Tanguay and Pierre Lavallée [1]

Abstract

To model economic depreciation, a database is used that contains information on assets discarded by companies. The acquisition and resale prices are known along with the length of use of these assets. However, the assets for which prices are known are only those that were involved in a transaction. While an asset depreciates on a continuous basis during its service life, the value of the asset is only known when there has been a transaction. This article proposes an *ex post* weighting to offset the effect of source of error in building econometric models.

Key Words: Price ratio; Survival data; Uniform distribution; Depreciation of vehicles.

1. Introduction

Various econometric models are used to estimate economic depreciation. To this end, we use a database containing information on assets discarded by companies. The acquisition and resale prices are known along with the length of use of these assets. From this information, we would like to infer results for the total population of assets used by companies. Regarding the use of the prices of used assets to estimate economic depreciation, we refer the reader to, Gellatly, Tanguay and Yan (2002) and Hulten and Wykoff (1981).

We question, however, the representativeness of the database used. Indeed, the assets for which prices are known are solely those subject to a transaction. We do not know the extent to which the losses of value observed on these assets are representative of the loss of value for all assets in production, regardless of whether they were the subject of a transaction. This situation can be a source of error in building econometric models because these models seek to measure depreciation of assets over their service lives, regardless of whether there was a transaction.

It is this second source of error that we propose to offset, at least in part, by applying *ex post* weighting when building econometric models. Section 2 of this article will describe the problem in greater detail, while in Section 3, we will describe the approach used to determine the weights. Finally, in Section 4, we present some numeric results.

2. Problem

We are seeking to describe the relationship between prices and asset age. There is a sample of n assets where we know, for each asset i, the price ratio r_i and the time t_i when this ratio was measured. Once prices are expressed in real dollars, this ratio is given as $r_i = P_i^t / P_i^0$ where P_i^0 is the initial value of the investment in asset i and P_i^t is its resale price at time t. This ratio is strictly decreasing in relation to the time axis t. At the start, we do not know the process that generates the loss in value and there are no specifics about the function that describes this loss except that it is strictly decreasing. However, it is possible to examine the distribution of the price ratios between 0 and 1. Here is an example constructed from data on manufacturing plants (note that 2/3 of the sample was excluded because it corresponds to discarded assets (the price is zero) and the estimation procedures take this component into account, each in its own way).

Since we want to use the data to infer statistics on the population of assets in production, we would like our data to have properties similar to those of a random sample drawn from that population. As we stated earlier, this is not the case because we only have the prices of assets i that were subject to a transaction at time t_i, $i = 1, ..., n$. In effect, while we would like to have price ratios for various periods in the existence of a given asset i, the ratio is only available when there has been a transaction, something that occurs in a non-uniform manner over an asset's service life.

Consequently, we can ask ourselves what form the above distribution might have if it had been drawn from a sample in which the price ratio had been measured, for the same asset i at different times t. Our argument is that it should converge toward a *uniform distribution*. We will therefore seek to obtain a weighting that will help us recreate a uniform distribution of price ratios. This weighting will help us offset the lack of uniformity in the distribution of observations, which may impact statistical analyses such as linear regression.

1. Marc Tanguay and Pierre Lavallée, Statistics Canada, Ottawa, Ontario, Canada K1A 0T6. E-mail: marc.tanguay@statcan.ca, pierre.lavallee@statcan.ca.

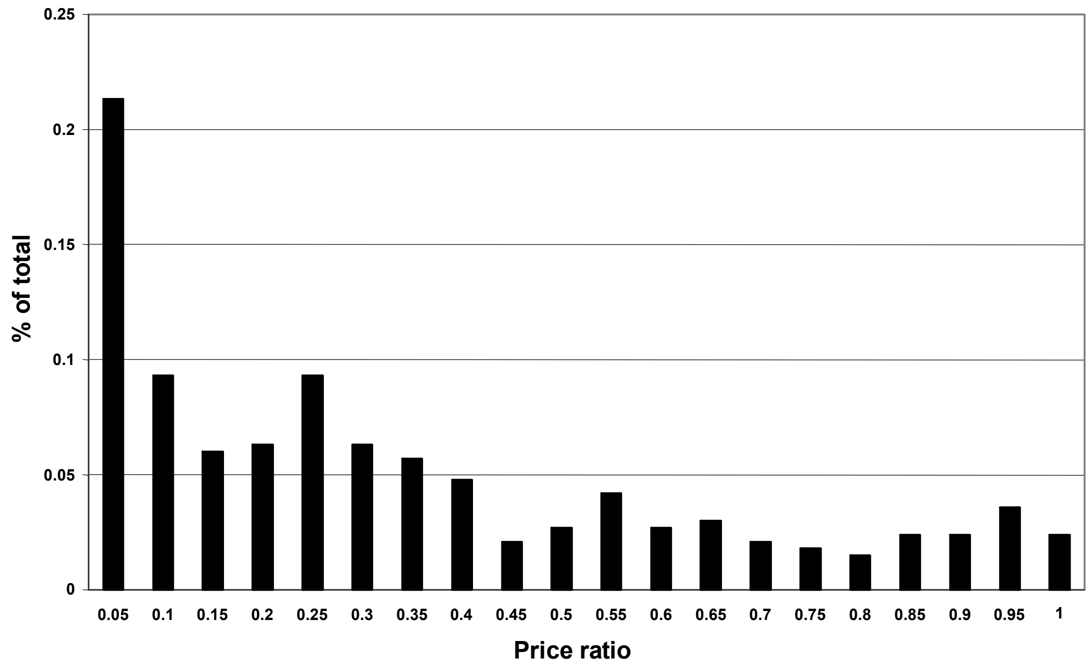

Figure 1 Distribution of observations by price ratio, manufacturing plants

3. Approach

Our starting point is that price ratios can be considered empirical realizations of an unknown form of survival function. In service life models, the survival function expresses the probability that an entity with a limited service life will survive beyond a certain point on the time axis. Accordingly, it provides the same information as a distribution function (or Cumulative Distribution Function). We will let $r,$ be a random variable describing the service life of a unit of value incorporated in some asset. The value gradually erodes over time for as long as the asset is in service. The price ratio can therefore be interpreted as the surviving fraction that gradually becomes smaller and smaller. This fraction is written as $S(y)$ and gives

$$S(y) = 1 - F(y)$$

where $F(y) = P(r \leq y)$ is the distribution function, that is, the probability that a unit of value is lost before point y.

Fundamental transformation theorems of probability laws provide the means for defining the inverse function of $F(y)$ (Greene 1993 and Ross 2002). We let $z = F(y)$ and assume that the inverse function F^{-1} exists so that $y = F^{-1}(z)$. This shows that there is a direct match between the space of y, bounded at 0 but infinite to the right, and that of F which is bound between 0 and 1. The distribution function of z is $F(F^{-1}(z)) = z$. The law that generates this distribution is a uniform distribution between 0 and 1.

This result is generally at the core of data generation processes like Monte Carlo simulations because the uniform distribution is often used when a random sample is being generated, followed by the application of the inverse function (Davidson and MacKinnon 1993). This approach is not always practical and indeed is sometimes patently impossible, especially if the inverse function F^{-1} is not explicit. This result has also been used in generalized remainder approaches, notably to build specification tests (Lancaster 1985).

The result is that any random sample built using empirical realizations of survival proportion data must converge in distribution toward a uniform distribution.

In the case of price data, intuition suggests that between the time of investment and that of disposal, the full range of relative prices must be covered by an asset in production. Initially, value depreciates faster and therefore there are more observations with short periods of time. This is offset by the fact that the corresponding reference on the time scale is also shorter. For example, it takes less time to move from 100% of the initial value to 90%, than from 15% to 5% of the initial value.

It is easy to verify these findings numerically using simulated data and we will not spend time on this. Rather, we will examine how this result can be reintroduced in the database to produce, at least partially, properties similar to those of a random sample. W*e can do this by simply imposing ex post on the empirical price distribution a* *weight structure* w_i *that ensures that the empirical distribution of the data, in the price space, is uniform.*

The empirical distribution of price ratios r is given by

$$\hat{F}_n(y) = \frac{\sum_{i=1}^n I_i(y)}{n} \quad (1)$$

where $I_i(y) = 1$ if the measured value r_i of asset i is less than or equal to y (specifically, $r_i \le y$), and 0 otherwise, and n is the total number of observations. Note that if the n units of the sample are independent and identically distributed (i.i.d.), when $n \to \infty$, $\hat{F}_n(y)$ converges in probability to $F(y)$, that is, $\hat{F}_n(y) \xrightarrow{P} F(y)$ (Bickel and Doksum 1977).

To obtain weight w_i for each asset i, we simply distribute the sample in a given number H of intervals (or classes) of a fixed size on the scale of price ratios, and we assign the same probability $\pi = 1/H$ to each of these intervals. Since the price ratios are bounded by 0 and 1, we then have the interval $h = 1$ given by $[0, H^{-1}]$, and for $h = 2, ..., H$, the intervals are given by $](h-1)H^{-1}, hH^{-1}]$. A weight w_h is then calculated in each interval h by the ratio $\pi / \hat{\pi}_h$ where $\hat{\pi}_h$ is the empirical probability specific to interval h, producing

$$\hat{\pi}_h = \frac{1}{n} \sum_{i=1}^{n} \delta_i(h) = \frac{n_h}{n} \quad (2)$$

where $\delta_i(h) = 1$ if $r_i \in h$, 0 otherwise. We then propose

$$w_i = w_h = \frac{\pi}{\hat{\pi}_h} = \frac{n}{Hn_h} \quad (3)$$

for $r_i \in h$. Using these weights, the *weighted empirical distribution* of the price ratios r is given by

$$\hat{F}_{n,w}(y) = \frac{\sum_{i=1}^n w_i I_i(y)}{\sum_{i=1}^n w_i}. \quad (4)$$

By writing $\sum_{i=1}^n w_i = \sum_{h=1}^H \sum_{i=1}^{n_h} n / Hn_h = n$, we finally get

$$\hat{F}_{n,w}(y) = \frac{\sum_{i=1}^n w_i I_i(y)}{n}. \quad (5)$$

Since $n_h = \sum_{i=1}^n \delta_i(h)$, we have

$$\begin{aligned}\hat{F}_{n,w}(y) &= \frac{\sum_{i=1}^n w_i I_i(y)}{n} \\ &= \frac{1}{H} \sum_{h=1}^{H} \frac{1}{n_h} \sum_{i=1}^{n} \delta_i(h) I_i(y) \\ &= \frac{1}{H} \sum_{h=1}^{H} \frac{\sum_{i=1}^n \delta_i(h) I_i(y)}{\sum_{i=1}^n \delta_i(h)} \\ &= \frac{1}{H} \sum_{h=1}^{H} \hat{F}_n(y \mid h). \quad (6)\end{aligned}$$

When $n \to \infty$, we have $(1/n) \sum_{i=1}^n \delta_i(h) I_i(y) \xrightarrow{P} P(r \in h,\, r \le y)$ and $(1/n) \sum_{i=1}^n \delta_i(h) \xrightarrow{P} P(r \in h)$. Thus, when $n \to \infty$,

$$\begin{aligned}\hat{F}_n(y \mid h) &\xrightarrow{P} \frac{P(r \in h, r \le y)}{P(r \in h)} \\ &= P(r \le y \mid r \in h) = F(y \mid h) \quad (7)\end{aligned}$$

where $F(y \mid h)$ is the distribution of price ratios r within interval h.

For a sufficiently large n, H must be determined in such a way as to build the intervals h so that $\hat{F}_n(y \mid h)$ is distributed approximately uniformly, $h = 1, ..., H$. In other words, when $n \to \infty$, for a sufficiently large H, $F(y \mid h)$ should have a uniform distribution on interval h. Note that this argument was used by Dalenius and Hodges (1959) in a context of optimal stratification. In this case, the distribution $F(y \mid h)$ is given by

$$F(y \mid h) = \begin{cases} 0 \text{ for } y \le (h-1)H^{-1} \\ Hy - h + 1 \text{ for } (h-1)H^{-1} < y \le hH^{-1} \\ 1 \text{ for } y > hH^{-1}. \end{cases} \quad (8)$$

Since $F(y) = \sum_{h=1}^H F(y \mid h)/H$, we have $F(y) = y$, which corresponds to the uniform distribution. We conclude from this that for a sufficiently large n, the use of weighting (3) should ensure that the weighted empirical distribution $\hat{F}_{n,w}(y)$ given by (5) is distributed approximately uniformly.

Monte Carlo simulations have shown that estimates produced from a non-random sample could be improved by using this approach. Its main advantages can be attributed to:

- its simplicity;
- the fact that it can be introduced *ex ante*, or prior to introducing the econometric model as such. Consequently, it does not require strong working hypotheses.

If we go back to the histogram presented earlier and divide the sample in $H = 5$ intervals of a width of 0.2 and a value of $\pi = 1/5 = 0.2$, we then get the following histogram that was weighted *ex post*.

4. Application

We will now illustrate our approach using an example taken from the Kelly Blue Book, a source of information widely used to estimate depreciation of automobiles. Table 1 shows the prices of two models of cars at different ages between 1 and 18 years. For each car, we have a sample of $n = 18$ units. Prices are expressed in relative value in relation to a new model. The ratios also have to be adjusted to take into account the survival probability at each of these ages. For each vehicle, the final ratio used r_i for year *i* is built from the product of the price ratio times the survival probability.

We are interested in the average depreciation rate $\overline{\tau}$ for each car. This can be estimated from a regression of the prices (or from a function of these prices) in relation to age (or a function of age). However, if we assume that the rate is constant and geometric, we obtain the relationship $r_i = 1-\overline{\tau}^i$, where r_i is the relative price based on age *i*. In this case, a rate $\hat{\tau}_i$ can be estimated at each age *i* by $\hat{\tau}_i = 1-r_i^{1/i}$. An estimate of the average rate of depreciation is then produced from the average for all ages, $\hat{\overline{\tau}} = \sum_{i=1}^{18} \hat{\tau}_i / 18$.

In the above example, we see that the depreciation rates $\hat{\tau}_i$ vary by age range and that they tend to increase with age. Moreover, the fact that we use a simple average of the ages in calculating $\hat{\overline{\tau}}$ again implicitly gives the same weight to each age. However, it is quite clear that this is not the distribution that we would get from a random sample of service vehicles. The figure below shows the distribution of price cells between ratios of 0 and 1.

The reweighting technique simply involves applying an equal weight to each of the relative price ranges. In this example, the $n = 18$ ages are distributed into $H = 7$ classes, resulting in 18/7 of the ages in each class (in reality, the structures of the cells was configured into 8 classes but the last is always empty). As mentioned in Section 3, the individual weights w_i for each age *i* are built using (3), that is, by dividing 18/7 by the number of observations found in each class, except for the empty cells where the weight remains zero. Table 2 shows the results and the impact of reweighting on the derived statistics.

This example clearly illustrates the problems of aggregation bias typical of regressions estimated from economic aggregates without taking account the real distribution of the units at the micro level. Thus, it is quite clear that the units at 17 and 18 years would not have the same regression weight as those at 1 year because the risk of loss at 1 year affects almost all vehicles to be put into circulation, while very few of them will be exposed to the risk of loss of value at more advanced ages. The result is that the unweighted estimate in this example produces an over-estimation of the depreciation rate in the order of 15%.

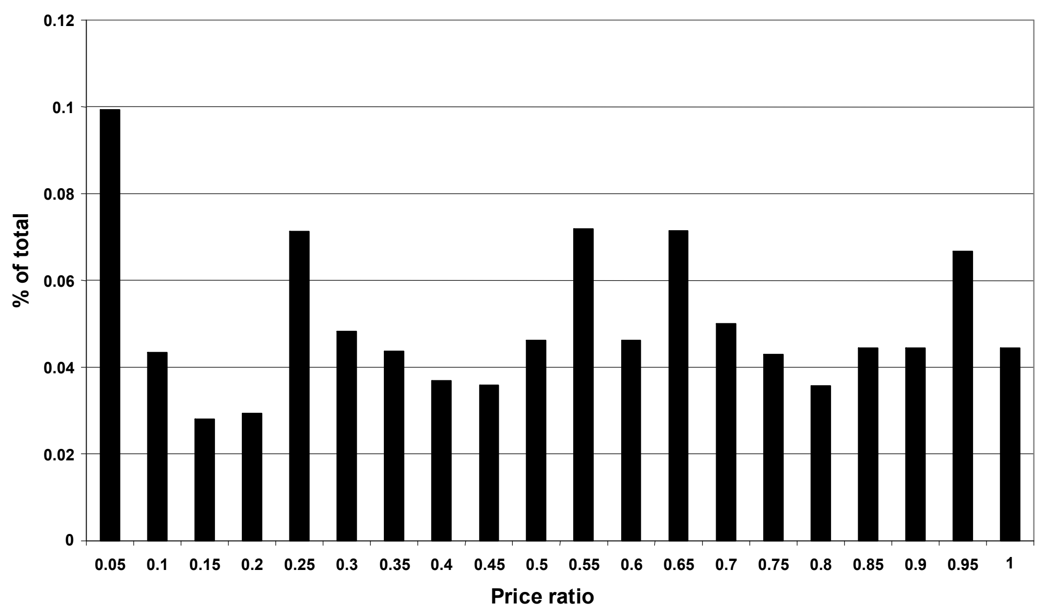

Figure 2 Weighted distribution of observations by price ratio, manufacturing plants *Ex post* weighting

Table 1 Relative prices of two models of cars based on the Kelly Blue Book and average depreciation rates before weighting

	Pr ($t > S$)*	Relative price				Average depreciation rates	
		Excluding disposals		Including disposals		Including disposals	
Year		*Buick*	*Chrysler*	*Buick*	*Chrysler*	*Buick*	*Chrysler*
1	0.9988	0.8633	0.8257	0.8622	0.8246	0.1367	0.1743
2	0.9901	0.7435	0.6801	0.7361	0.6734	0.1377	0.1753
3	0.9666	0.6410	0.5608	0.6195	0.5420	0.1378	0.1754
4	0.9220	0.5523	0.4621	0.5092	0.4261	0.1379	0.1755
5	0.8526	0.4740	0.3794	0.4042	0.3234	0.1387	0.1762
6	0.7582	0.4034	0.3087	0.3058	0.2341	0.1404	0.1779
7	0.6433	0.3391	0.2482	0.2181	0.1597	0.1432	0.1805
8	0.5164	0.2790	0.1953	0.1441	0.1009	0.1475	0.1846
9	0.3892	0.2227	0.1491	0.0867	0.0580	0.1537	0.1906
10	0.2731	0.1639	0.1050	0.0448	0.0287	0.1654	0.2018
11	0.1770	0.1261	0.0772	0.0223	0.0137	0.1716	0.2077
12	0.1051	0.0892	0.0523	0.0094	0.0055	0.1824	0.2180
13	0.0567	0.0614	0.0344	0.0035	0.0019	0.1932	0.2284
14	0.0276	0.0441	0.0236	0.0012	0.0007	0.1999	0.2347
15	0.0120	0.0320	0.0164	0.0004	0.0002	0.2050	0.2396
16	0.0046	0.0190	0.0093	0.0001	0.0000	0.2194	0.2534
17	0.0016	0.0088	0.0041	0.0000	0.0000	0.2432	0.2761
18	0.0005	0.0051	0.0023	0.0000	0.0000	0.2542	0.2867
					Average	0.1727	0.2087

*Survival probability based on estimates from the Micro-Economic Studies and Analysis Division of Statistics Canada.

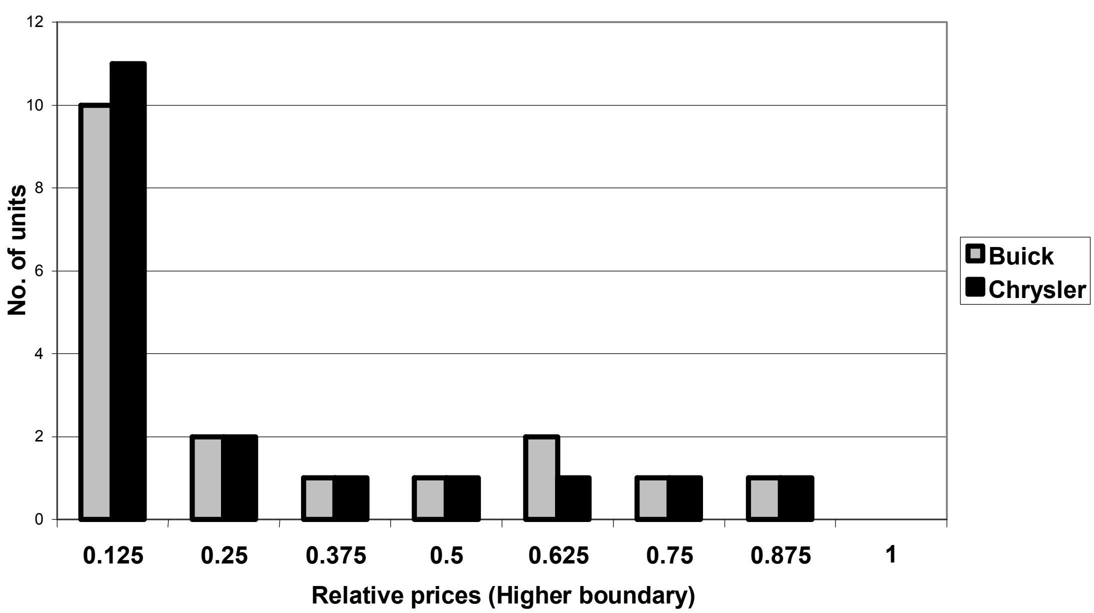

Figure 3 Distribution of cells used to estimate the average depreciation rate using data from the Kelly Blue Book before weighting (Total = 18)

Table 2 Relative prices of two models of cars based on the Kelly Blue Book and the average depreciation rate after weighting

	Relative prices		**Average depreciation rates**		***Ex post* weights**	
	Including disposals		Including disposals			
Year	*Buick*	*Chrysler*	*Buick*	*Chrysler*	*Buick*	*Chrysler*
1	0.8622	0.8246	0.1367	0.1743	2.5714	2.5714
2	0.7361	0.6734	0.1377	0.1753	2.5714	2.5714
3	0.6195	0.5420	0.1378	0.1754	1.2857	2.5714
4	0.5092	0.4261	0.1379	0.1755	1.2857	2.5714
5	0.4042	0.3234	0.1387	0.1762	2.5714	2.5714
6	0.3058	0.2341	0.1404	0.1779	2.5714	1.2857
7	0.2181	0.1597	0.1432	0.1805	1.2857	1.2857
8	0.1441	0.1009	0.1475	0.1846	1.2857	0.2338
9	0.0867	0.0580	0.1537	0.1906	0.2571	0.2338
10	0.0448	0.0287	0.1654	0.2018	0.2571	0.2338
11	0.0223	0.0137	0.1716	0.2077	0.2571	0.2338
12	0.0094	0.0055	0.1824	0.2180	0.2571	0.2338
13	0.0035	0.0019	0.1932	0.2284	0.2571	0.2338
14	0.0012	0.0007	0.1999	0.2347	0.2571	0.2338
15	0.0004	0.0002	0.2050	0.2396	0.2571	0.2338
16	0.0001	0.0000	0.2194	0.2534	0.2571	0.2338
17	0.0000	0.0000	0.2432	0.2761	0.2571	0.2338
18	0.0000	0.0000	0.2542	0.2867	0.2571	0.2338
				Weighted average	0.1479	0.1836

Acknowledgements

The authors would like to express their sincere appreciation to the anonymous referee of *Survey Methodology*, whose thoughtful comments helped improve the quality of the article.

References

Bickel, P.J., and Doksum, K.A. (1977). *Mathematical Statistics*, Holden-Day, Oakland, CA.

Dalenius, T., and Hodges, J.L. (1959). Minimum variance stratification. *Journal of the American Statistical Association*, 54, 88-101.

Davidson, R., and MacKinnon, J.G. (1993). *Estimation and Inference in Econometrics*. Oxford University Press, N.Y.

Gellatly, G., Tanguay, M. and Yan, B. (2002). An alternative methodology for estimating economic depreciation: New results using a survival model. In *Productivity Growth in Canada - 2002*, Statistics Canada. #15-204-XPE.

Greene, W.H. (1993). *Econometric Analysis*. Second edition, Prentice Hall, Englewood Cliffs, N.J.

Hulten, C.R., and Wykoff, F.C. (1981). The measurement of economic depreciation. In *Depreciation, Inflation, and the Taxation of Income from Capital*, (Ed. C.R. Hulten). The Urban Institute Press, Washington, D.C, 81-125.

Lancaster, T. (1985). Generalized residuals and heterogeneous duration model: With applications to the weibull model. *Journal of Econometrics*, 28, 155-69.

Ross, S.M. (2002). *Introduction to Probability Models*, 8st Edition, Academic Press.