Journal of the Korean Institute of Industrial Engineers
Vol. 40, No. 3, pp. 283-290, June 2014.
ISSN 1225-0988 | EISSN 2234-6457
http://dx.doi.org/10.7232/JKIIE.2014.40.3.283
© 2014 KIIE
<Original Research Paper>

생존분석을 이용한 디스플레이 FAB의 반송시간 예측모형

한바울 · 백준걸†

고려대학교 산업경영공학과

Prediction Model on Delivery Time in Display FAB Using Survival Analysis

Paul Han · Jun Geol Baek

School of Industrial Management Engineering, Korea University

In the flat panel display industry, to meet production target quantities and the deadline of production, the scheduler and dispatching systems are major production management systems which control the order of facility production and the distribution of WIP (Work In Process). Especially the delivery time is a key factor of the dispatching system for the time when a lot can be supplied to the facility. In this paper, we use survival analysis methods to identify main factors of the delivery time and to build the delivery time forecasting model. To select important explanatory variables, the cox proportional hazard model is used to. To make a prediction model, the accelerated failure time (AFT) model was used. Performance comparisons were conducted with two other models, which are the technical statistics model based on transfer history and the linear regression model using same explanatory variables with AFT model. As a result, the mean square error (MSE) criteria, the AFT model decreased by 33.8% compared to the statistics prediction model, decreased by 5.3% compared to the linear regression model. This survival analysis approach is applicable to implementing the delivery time estimator in display manufacturing. And it can contribute to improve the productivity and reliability of production management system.

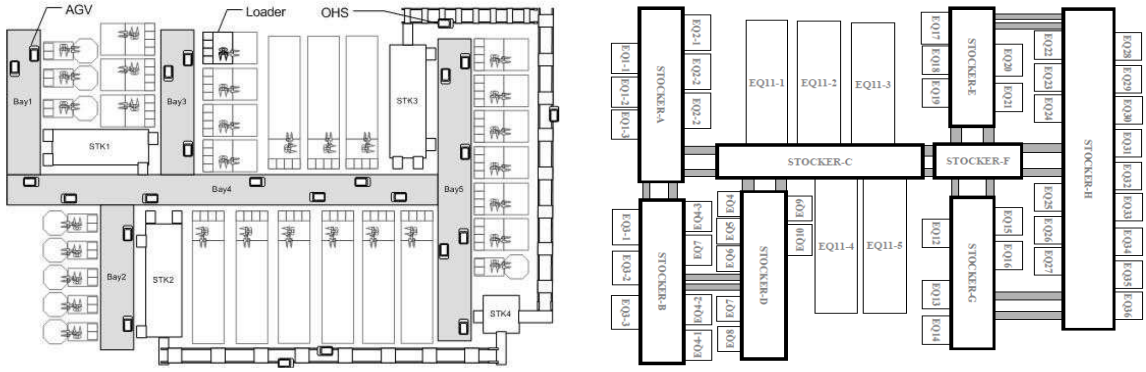
Keywords: Delivery Time, Survival Analysis, Cox PH Model, Accelerated Failure Time Model

1. 서 론

1.1 디스플레이 FAB 구조와 반송 특징

첨단 산업을 대표하는 디스플레이 제조공장에서 납기준수와 생산량을 극대화하기 위해 스케줄러(Scheduler)와 디스패칭(Dispatching) 시스템에 대한 품질향상 요구가 커지고 있다. 주요 재료인 원판유리(Mother glass) 크기가 증가하면서 적재 공간의 제약을 많이 받게 되었고, 이로 인해 적정 재공재고(Work in process; WIP)에 대한 관리와 제어가 반도체 라인과 비교하여 중요한 관리요소가 되고 있다. 반도체와 디스플레이 산업의 공통적인 생산 특성은 한 설비에서 공정 조건을 변경하며 여러 공정을 진행하는 잡샵(Job Shop)과 다양한 프로세스의 제품들을 동시에 진행하는 병렬기계 생산(Parallel machine) 구조이다. 이와 같은 특성을 고려하여 라인설계는 최적의 물류동선과 생산변화에 유연한 레이아웃을 추구하고 있다. 1990년대 말까지 디스플레이 FAB(Fabrication)은 반도체 FAB와 유사한 AGV(Auto guided vehicle)가 주요 물류설비였으나, 원판유리가 대형화되면서 효율적인 공간사용을 위해 인라인 스토커(Inline Stocker, 이후 스토커로 명칭)가 주요 물류설비가 되었다(Gu

본 연구는 지식경제부 및 정보통신산업진흥원의 정보통신연구기반구축사업의 연구결과로 수행되었음(NIPA-2013-(I2218-13-1004)). 이 논문은 2013년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(NRF-2013R1A1A2010019).
† 연락저자 : 백준걸 교수, 136-701 서울시 성북구 안암동 5가 1번지 고려대학교 산업경영공학과, Tel : 02-3290-3396, Fax : 02-929-5888, E-mail : jungeol@korea.ac.kr
2013년 12월 30일 접수; 2014년 3월 5일 수정본 접수; 2014년 3월 27일 게재 확정.



AGV layout Stocker layout

Figure 1. AGV layout vs. Stocker layout

and Choi, 2008). 스토커는 공정설비와 직접 연결되어 재공을 공급, 회수하는 역할과 버퍼역할을 동시에 수행한다(Jang and Choi, 2006; Chung and Kim, 2008). <Figure 1>은 AGV 중심의 레이아웃과 스토커 중심의 레이아웃을 비교한 그림이다(Son and Choi, 2008).

디스플레이 FAB에서 원판유리의 설비 간 이동은 카세트(Cassette)라는 컨테이너에 담겨 발생하기 때문에 카세트가 반송의 기본단위가 된다(Chung *et al*., 2010).

따라서 본 논문에서 다루는 반송은 카세트를 정해진 목적지까지 운반하는 것을 의미한다. 반송은 매 공정마다 2회 정도 발생하는데, <Figure 2>는 원판유리가 한 공정을 진행하는데 소요되는 시간(단위공정 공기)을 세분화한 그림이다.

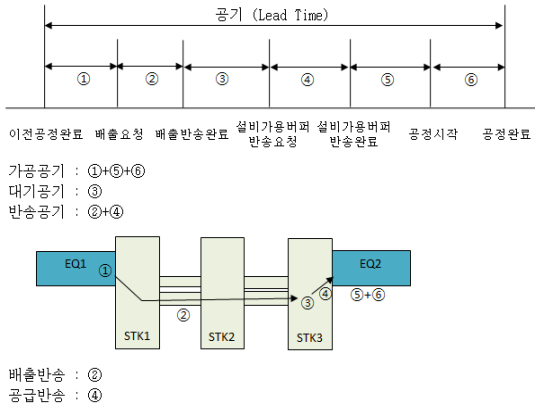


Figure 2. Detailed lead time of one production step

<Figure 2>에서 ②번 반송은 설비에서 카세트가 배출될 때 발생하는 배출반송이고 ④번 반송은 다음 공정설비에서 재공을 요청할 때 발생하는 공급반송이다. 정상적인 Lot이 들어있는 카세트는 이전 설비에서 배출되기 전에 디스패칭 시스템에 의해 이미 후보 설비군 중 다음 진행 설비가 지정된다. 따라서 배출반송 시 다음 공정설비와 인접한 스토커를 목적지로 정해 반송이 실행되기 때문에 배출반송(②번)이 공급반송(④번)보다 이동거리와 반송시간이 긴 것을 알 수 있다.

스케줄러와 디스패칭 시스템은 생산량과 출하납기를 맞추기 위해 공정단위로 납기시간을 계획한다. 최근의 생산 환경을 보면 높은 설비 안정성으로 가공공기는 낮은 편차를 보이는 반면, 반송공기는 생산량 증가와 병렬 및 혼류생산으로 편차가 증가하고 있다. 또한 공장 내 재공이 감소하면서 대기공기가 줄어들어 상대적으로 반송공기가 차지하는 비중이 증가하고 있다. 따라서 반송시간 예측의 정확성은 스케줄러와 디스패칭 시스템 품질에 필수요소가 된다(Shabtay, 2010).

1.2 기존 연구

기존에 반송시간을 예측하는 모형은 크게 3가지로 분류할 수 있다.

- 기술 통계(Descriptive statistics) : 과거의 반송 데이터를 바탕으로 통계량(평균 및 분산)을 사용하여 반송시간을 예측하는 방법이며 기존 시스템에서 사용하는 방식이다.
- 통계적 분석 모형(Analytical model) : 반송시간에 영향을 주는 변수들을 도출하여 다변량(Multivariate) 자료 분석 기법을 이용하는 방법이다.
- 시뮬레이션 모형(Simulation model) : 중요변수를 도출하여 변수들의 분포와 반송 로직을 모델링하여 실제 상황과 유사한 모형을 만들어 시뮬레이션을 통해 반송시간을 예측하는 방법이다(Kong, 2007; DeJong and Wu, 2002).

지금까지 통계적 분석모형(Analytical model)에 대한 연구는 다변량 회귀분석과 가중이동평균(Lee *et al*., 2009)을 이용한 접근이 많았으며, 최근에는 신경망 모형(Neural network model)(Liao and Wang, 2004; Chen, 2009), SVM(Support vector machine)(Shiue, 2009) 등의 데이터마이닝 기법을 이용한 연구들이 활발하게 이루어지고 있다.

본 연구에서는 디스플레이 라인의 반송시간과 생존분석의 생존시간 분포의 유사성을 이용하여 반송시간을 추정한다(Smith, 1975). 생존분석기법 중에 비모수적 다변량 통계기법인 Cox 비례위험 모형(Cox, 1972)과 모수적 기법인 가속화 고장시간 모형(Accelerated failure time, 이후 AFT 모형으로 명칭)(Wei, 1992)

을 활용하여 반송시간 예측 모형을 제안하고자 한다.

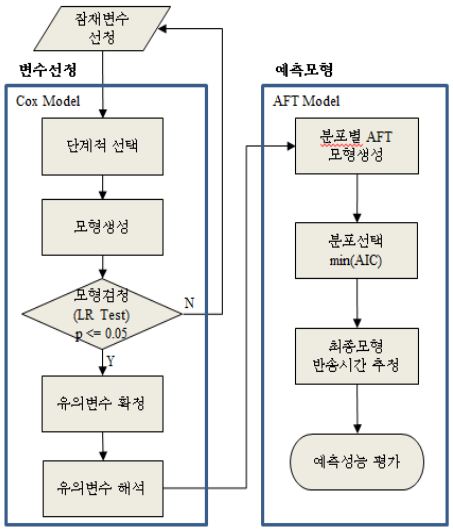


Figure 3. Process of prediction model using the survival analysis

<Figure 3>은 변수 선택에서 모형의 성능 평가까지 일련의 과정을 요약한 그림이다. 제 3장에서 Cox 비례위험 모형을 통해 변수선정 과정을 설명하며, 제 4장에서 선정된 변수를 기반으로 AFT 반송예측 모형을 구축하여 성능평가를 진행한다. 제 5장 결론에서는 본 연구의 가치와 추후 연구 과제를 제시한다.

2. 생존분석을 이용한 시간예측 접근

2.1 개 요

반송시간과 같이 사건의 시작부터 종료까지 사건의 지속시간을 의미하는 데이터(Time to event)는 0 미만 값을 가질 수 없기 때문에 편향된 분포의 특성을 가지고 있다(Hosmer *et al*., 2011). 본 연구를 위해 수집된 디스플레이 공장의 반송시간 자료를 선형회귀 모형으로 만들면 잔차(residual)의 분포가 <Figure 4>과 같이 정규분포를 따르지 않음을 알 수 있다.

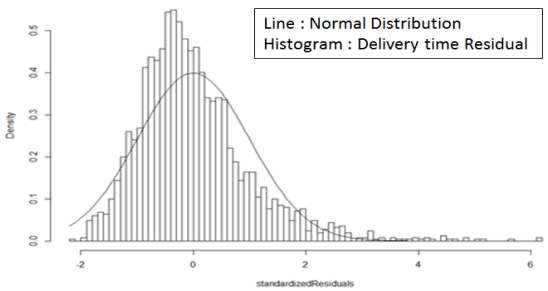


Figure 4. Distribution of delivery time residual

선형회귀 모형으로는 종속변수인 반송시간의 분포를 충분히 반영하지 못하기 때문에 모형 추정치에 편의(bias)가 발생하고, 오차항에도 편의된(biased) 요소가 존재하게 된다.

이와 같이 추정치의 정확한 분포를 판단하기 어려운 경우 예측이나 분류의 정확성을 높이기 위해 일반화 선형 혼합모형(generalized linear mixed model)이나 비선형 회귀모형으로 접근하는 경우가 많다(McCulloch, 2003). 반면 생존분석 기법 중 Cox 비례모형방법은 생존시간(반송시간)에 대한 분포 가정 없이 모형에 접근할 수 있고, 예측 결과와 더불어 중요변수들과 반송시간의 관계 해석이 용이하다는 장점이 있다(Cox, 1972). 또한 모수적 생존분석 기법인 AFT 모형을 이용하여 가장 근접한 분포를 찾아내고 생존시간을 추정할 수 있다(Jenkins, 2005; Qi, 2009).

2.2 생존 분석 기법

이 절에서는 생존분석의 기본성질을 먼저 논의하고 Cox 비례모형과 AFT 모형의 특성을 알아본다. 반송시간처럼 특정 사건(Event)의 발생부터 종료 시점까지의 지속시간을 추정하거나 분석하는 기법을 생존분석이라 하며 시간분석(Time to event analysis)이라고도 불린다(Cox, 1972; Fox, 2002). 생존분석은 관심의 대상인 모집단의 생존시간에 대한 통계적 추론을 목적으로 하며, 의학과 임상 분야에서 널리 사용되고 있다(Hosmer *et al*., 2011). 사건이 발생한 시점을 t_0, 사건이 종료된 시점을 T라고 했을 때, $t = T - t_0$를 생존기간으로 정의한다. 반송 관점에서 $t_0 = 0$으로 가정하면 반송의 생존기간, 즉 반송시간은 식 (1)과 같이 확률밀도함수 $f(t)$를 갖는 확률변수가 된다.

$$f(t) = \lim_{\Delta t \to 0} \frac{\Pr(t \le T \le t + \Delta t)}{\Delta t} \tag{1}$$

밀도함수 $f(t)$로부터 확률 누적함수$F(t)$를 식 (2)와 같이 정의하면

$$F(t) = \Pr(T \le t) = \int_0^t f(u)du \tag{2}$$

반송이 t시점까지 종료되지 않을 확률을 의미하는 생존확률함수 $S(t)$는 식 (3)과 같이 표현된다.

$$S(t) = \Pr(T > t) = \int_t^\infty f(u)du = 1 - F(t) \tag{3}$$

식 (2)와 식 (3)으로부터 식 (4)를 도출할 수 있다.

$$f(t) = F'(t) = -S'(t) \tag{4}$$

위험의 정도, 여기에서는 반송이 종료될 확률을 의미하는

위험함수 $h(t)$는 t시점까지 생존하에 순간조건부 확률의 극한치이므로 식 (5)과 같다.

$$h(t)=\lim_{\Delta t\to 0}\frac{\Pr(t\le T\le t+\Delta t|\, T>t)}{\Delta t},\ t\ge 0 \quad (5)$$

위험함수 $h(t)$를 확률밀도함수 $f(t)$와 생존함수 $S(t)$로 표현하면 식 (6) 같이 전개할 수 있으며,

$$h(t)=\frac{f(t)}{S(t)}=\frac{-\frac{d}{dt}S(t)}{S(t)}=-\frac{d}{dt}logS(t) \quad (6)$$

생존함수 $S(t)$와 위험함수 $h(t)$간의 관계를 식 (7)과 같이 정리할 수 있고, 음의 상관관계가 있다.

$$S(t)=\exp\left[-\int_0^t h(u)du\right] \quad (7)$$

Cox는 위험함수 $h(t)$를 식 (8)과 같이 정의를 하였다(Cox, 1972).

$$h(t,\, X)=h_0(t)\exp(\beta_1x_1+\beta_2x_2+\cdots+\beta_kx_k) \quad (8)$$

$h(t,\, X)$는 설명변수 $x_1,\, x_2,\, \cdots,\, x_k$가 주어졌을 때 시간 t에서 반송이 종료될 위험함수이고, $h_0(t)$는 설명변수와 상관없이 시간 t에 의존하는 값이며 이를 기저위험 함수(baseline hazard function)라고 한다(Cox, 1972). 여기서 $\beta_1\cdots\beta_k$는 추정해야 할 대상이 되며 한 단위 증가할 때의 상대적 위험비(relative risk)를 의미한다. Cox 모형은 기저위험 함수가 명확한 형태를 보이지 않고, 설명변수들의 영향도로 종속변수인 위험함수가 표현되기 때문에 비모수적 추정방법이다. 따라서 설명변수와 생존기간(반송시간) 사이의 영향도 분석에는 용이하지만 생존기간을 예측하기에 적절하지 않다.

Table 1. Survival and hazard function of distributions

Distribution	$f(t)$	$S(t)$	$h(t)$
Exponential	$\lambda\exp(-\lambda t)$	$\exp(-\lambda t)$	λ
Weibull	$\lambda pt^{p-1}\exp(-\lambda t^p)$	$\exp(-\lambda t^p)$	λpt^{p-1}
Log-logistic	$\frac{\lambda pt^{p-1}}{(1+\lambda t^p)^2}$	$\frac{1}{1+\lambda t^p}$	$\frac{\lambda pt^{p-1}}{1+\lambda t^p}$

$\lambda=\exp(\beta^Tx)$.

생존기간을 예측하기 위해서는 Cox 비례위험 모형의 모수적 기법인 AFT 모형을 사용한다(Lin and Wei, 1992). 모수적 방법을 적용하기 위해서는 Cox 비례위험 모형의 기저위험 함수 $h_0(t)$에 대한 특정 형태의 함수적 모형이나 분포를 가진다는 가정이 필요한데, <Table 1>은 생존분석에서 주로 사용되는 분포들의 확률변수$f(t)$, 생존함수$S(t)$, 위험함수 $h(t)$를 나타낸다(Jenkins, 2005).

λ는 추정하는 회귀식을 지수함수로 표현한 인자이다. 회귀계수 β^T는 추정을 위해 최대우도추정(maximum likelihood estimate;MLE)을 사용하며, 생존시간 추정치는 생존함수의 중앙값 $S(t)=0.5$에서의 시간 $\hat{t_m}$이다(Qi, 2009). 예를 들어 지수분포에서 생존시간 추정식은 식 (9)와 같다.

$$S(t)=0.5\ \Rightarrow\ \hat{t_m}=\frac{-\log(0.5)}{\lambda}=\frac{-\log(0.5)}{\exp(\beta^Tx)} \quad (9)$$

AFT 모형은 설명변수의 다중 회귀식에 의한 생존시간 예측 모형은 가능하지만 모수적 분포가정의 한계를 가지고 있다. 따라서 본 연구에서는 Cox 비례위험 모형을 이용하여 중요설명변수를 선택하고, 주요 분포에 대해 각각 AFT 모형을 만들어 AIC(Akaike Information Criterion)값을 기준으로 예측성능을 비교하여 최종 모형을 선정한다. AIC에 대한 상세 내용은 4장에서 설명한다.

3. Cox 모형을 이용한 중요변수선정

3.1 분석자료

본 연구에서는 모 디스플레이 회사의 주요 생산 공정 중 연속된 3개 공정의 반송 이력을 표본 대상으로 선정하였다. 표본은 총 6,577개(경로 1 : 2,498, 경로 2 : 1,581, 경로 3 : 2,498)이며, 예측 모형을 만들기 위한 훈련 데이터와 예측 성능 평가를 위한 테스트 데이터를 1 : 1로 동일하게 나누었다. 3개의 표본 반송경로 형태는 <Figure 5>의 레이아웃과 같으며, 반송 설비의 돌발 정지, 유지보수로 인한 다른 우회경로 반송, 반송 시간이 60분을 초과한 이상치는 제외하였다. 경로 1은 카세트를 저장하는 선반의 개수가 출발스토커와 도착스토커 간 큰 차이를 보이며, 경로 2는 층간 카세트 반송이 발생하는 특징이 있고, 경로 3은 단순한 단방향 컨베이어로 연결되는 비슷한 크기의 스토커로 구성된 일반적인 반송 경로의 특징을 가지고 있다.

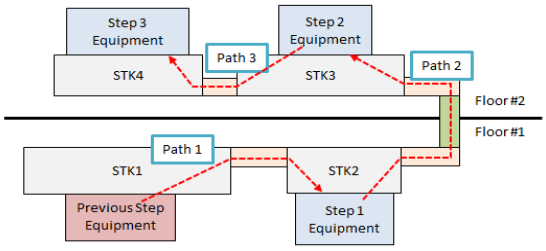


Figure 5. Experiment line layout and delivery path

모든 물류설비가 사양과 구조가 상이하기 때문에 반송시간

Table 2. Definition of variables

변수명	설명	데이터 유형
반송 긴급도(X1)	디스패칭 시스템에서 설정되는 반송의 긴급정도 (높을수록 긴급도가 높음)	정수(3, 4, …, 9)
출발지스토커 대기반송(X2)	출발지 스토커에 반송 실행 전 대기 상태인 반송수	정수(0, 1, 2, …)
목적지스토커 대기반송(X3)	목적지 스토커에 반송 실행 전 대기 상태인 반송수	정수(0, 1, 2, …)
출발지스토커 저장률(X4)	출발지 스토커 저장가능 공간 중 카세트가 저장되어 있는 비율 (카세트 수/저장 가능 선반수)×100	정수(0, 1, …, 100)
목적지스토커 저장률(X5)	목적지 스토커 저장가능 공간 중 카세트가 저장되어 있는 비율 (카세트 수/저장 가능 선반수)×100	정수(0, 1, …, 100)
컨베이어 카세트(X6)	컨베이어에 있는 카세트 수 (컨베이어 : 스토커를 연결하는 단방향 물류 설비)	정수(0, 1, 2, …)
Glass 저장매수(X7)	카세트에 담겨있는 Glass 매수	정수(0, 1, …, 20)
잔여 반송 납기시간(X8)	현재부터 스케줄러에서 진행 시간을 예측한 남은 시간 (반송 납기시간-현재시간)	연속형
컨베이어 대기반송수(X9)	출발지스토커 대기 반송 중 단위목적지가 컨베이어인 반송수	정수(0, 1, 2, …)
출발지스토커 반송순위(X10)	출발지스토커 반송 실행 순위 기준 1 : 긴급도 높을수록 기준 2 : 긴급도가 같을 경우 대기시간이 길수록	정수(1, 2, …)
목적지스토커 반송순위(X11)	목적지스토커 반송 실행 순위 기준 1 : 긴급도 높을수록 기준 2 : 긴급도가 같을 경우 대기시간이 길수록	정수(1, 2, …)
최근 평균 반송소요시간(X12)	최근 완료된 5개 동일한 반송 경로와 긴급도를 가진 반송들의 평균 반송 소요시간	연속형

의 중요 인자와 영향도가 경로마다 다르다. 따라서 모형의 예측성능을 높이기 위해 반송경로 단위로 모형을 만들었다. 모형을 추정하기 위한 후보변수는 반송이 발생하는 순간에 수집되는 물류상황에 대한 정보들이며 물류제어시스템(Automatic material handling system; AMHS)에서 계산하여 데이터베이스(Database)에 저장된다.

후보변수는 전문가 의견을 반영하여 총 12개를 선정하였다. 변수 정의는 <Table 2>와 같다. 모든 변수들은 정량적 변수(Quantitative variable)이며 정성적 변수(Qualitative variable) 표현을 위한 더미변수(Dummy variable)는 사용되지 않았다.

3.2 중요변수 선정

Cox 비례위험 모형에서 중요변수를 추출하기 위해 평가기준으로 AIC(Akaike Information Criterion)를 적용하는 단계적 선택법(Stepwise method)을 사용하였다. <Table 3>은 경로별로 선택한 변수의 유의수준 p-value와 위험비를 나타낸다.

위험비가 1 이상이면 변수의 값이 증가할 때마다 위험비가 증가하는 것으로, 반송시간 모형에서는 반송시간이 줄어듦을 의미한다. 모든 경로에 대해 반송긴급도(X1), Glass 저장매수(X7), 잔여반송납기 시간(X8)이 공통적으로 변수선정이 되었다. 특히 위험비를 통해 반송긴급도(X1)가 반송시간에 가장 큰 영향력을 주는 변수임을 알 수 있다. 경로 1에서 반송긴급도(X1)의 위험비는 1.16로 반송긴급도가 한 단계 상승할 때마다 반송시간이 16%씩 감소함을 의미한다. 실제 스토커에서 이동작업의 순서를 계획하는 스케줄러는 대기하고 있는 반송 중에 다음 반송을 선정할 때 반송 긴급도를 높은 우선순위로 사용하고 있다. 경로 2에서 특별히 주목할 변수는 위험비가 0.89인

Table 3. Variable selection in each paths

	경로 1		경로 2		경로 3	
변수	p	위험비	p	위험비	p	위험비
X1	0.00	1.16	0.00	1.12	0.04	1.06
X2	0.00	0.97	-	-	-	-
X3	-	-	0.00	0.97	0.00	0.97
X4	-	-	0.04	0.99	0.00	0.99
X5	0.05	0.99	-	-	-	-
X6	-	-	-	-	0.09	0.95
X7	0.03	0.98	0.01	0.97	0.01	0.98
X8	0.00	0.99	0.00	0.99	0.01	0.99
X9	-	-	0.00	0.89	0.00	0.96
X10	-	-	0.00	0.93	0.00	0.97
X11	-	-	0.02	0.96	-	-
X12	0.03	0.94	-	-	0.01	0.93

컨베이어 대기 반송수(X9)이다. 이는 스토커를 연결하는 컨베이어로 나가는 반송이 1개 증가할 때마다 반송시간이 11%씩 증가함을 의미하는데, 경로 2 상에 설치된 컨베이어는 컨베이어 중간지점에 승강기가 있는 복잡한 설계구조로 되어있다. 승강기로 인해 카세트 상하이동이 추가적으로 발생하면서 다른 컨베이어 비해 현저하게 낮은 수송능력을 나타낸다. 따라서 해당 컨베이어를 경유하는 반송은 스토커에서 대기하는 시간이 다른 반송들보다 길게 나타남을 위험비를 통해 알 수 있다.

Cox 비례위험 모형의 타당성을 검정하기 위해 LR(Likelihood ratio) 테스트를 사용하는데, 추정된 모수에 대한 가설과 검정식에 대해 변수를 포함하는 모형(alternative)이 포함하지 않는 모형(null)에 비해 몇 배 더 적합한가를 표현하는 방법이다. 테스트 결과는 <Table 4>와 같이 3개 경로 모두 1% 유의수준에서 유의한 것으로 나타났기 때문에 변수를 포함하는 것이 의미있는 것임을 알 수 있다.

Table 4. LR Test of Cox regression

	경로 1	경로 2	경로 3
Likelihood Ratio	540	223	319
p-value	< 0.001	< 0.001	< 0.001

4. AFT 모형을 이용한 반송시간 예측

4.1 모형 선정

제 3장에서 Cox 모형을 통해 경로별로 중요변수를 선택하였다. Cox 모형의 주목적은 반송시간의 추정이 아닌 중요변수의 효과를 추정하기 위함이며, 반송시간을 예측하기 위해서 AFT 모형을 사용한다. AFT 모형과 같은 모수적 생존분석기법은 생존함수에 대한 적절한 분포의 가정이 선결되어야 하는데, 많이 사용되는 분포로는 지수, 와이블, 로그정규, 로그로지스틱 분포 등이 있다. Bradburn *et al.*(2003)은 AFT 모형의 적합한 분포에 대해 식 (10)과 같이 AIC 척도를 사용하여 AIC 값이 작을수록 생존함수에 대한 설명력이 높은 분포임을 제안하였다.

$$AIC = -2Log(L) + 2(k + c + 1) \quad (10)$$

L : 우도$(likelihood)$

k : 변수의 수

c : 분포 파라미터

지수분포$(c = 1)$

와이블, 로그정규, 로그로지스틱$(c = 2)$

대상이 되는 3개 경로에 대해 <Table 5>와 같이 4가지 분포의 AIC 값을 비교하였으며, 모든 경로에서 로그정규 분포모형의 AIC 값이 가장 작았다. 따라서 본 연구에서는 로그정규 분포를 사용하여 반송시간 예측 모형을 구축하였다.

Table 5. AIC each models

분포	경로 1	경로 2	경로 3
exponential	8885.4	5040.2	8557.1
Weibull	7539.9	4227.3	7175.0
log-logistic	7366.1	4105.7	7072.2
log-normal	**7353.0**	**4092.6**	**7058.7**

반송시간을 추정하는 로그정규분포 AFT 모형의 회귀식은 식 (11)~식 (13)과 같다.

경로 1

$$T = \exp(1.697 + (-0.043)x_1 + 0.022x_2 + 0.002x_5 + 0.008x_7 + 0.003x_8 + 0.031x_{12}) \quad (11)$$

경로 2

$$T = \exp(1.352 + (-0.042)x_1 + 0.012x_3 + 0.005x_4 + 0.010x_7 + 0.010x_8 + 0.041x_9 + 0.033x_{10} + 0.016x_{11}) \quad (12)$$

경로 3

$$T = \exp(1.422 + (-0.029)x_1 + 0.014x_3 + 0.004x_4 + 0.040x_6 + 0.007x_7 + 0.003x_8 + 0.019x_9 + 0.013x_{10} + 0.019x_{12}) \quad (13)$$

4.2 반송시간 추정

만들어진 모형의 예측성능 검증은 테스트 데이터를 사용하여 진행한다. 반송시간 예측모형의 비교 대상은 대상 기업에서 현재 사용 중인 긴급도별 이력 평균시간으로 만들어진 모형(이후 평균 모형이라고 명칭)과 Cox 모형에서 선정한 동일 설명변수를 사용한 선형회귀 모형이다. 평균 모형은 과거 하루치의 반송이력을 8시간씩 나누어 그중 4분위수에 해당하는 데이터는 제외하여 평균값들을 구한 후 현재와 가까울수록 가중치를 높게 설정하여 계산하는 가중평균 방식이다. 4분위수를 제외시키는 이유는 이상치를 제거하기 위함이다.

모형의 예측성능 평가척도로는 평균제곱오차(mean squared error; MSE)를 사용하였다. 평균제곱오차 수식은 식 (14)과 같으며 L은 테스트 표본을 의미한다.

$$MSE_{test} = \frac{1}{n}\sum_{y_i \in L}(y_i - \hat{y_i})^2 \quad (14)$$

AFT 모형의 성능검정은 단측 t검정으로 판단한다. 귀무가설(H_0)은 'AFT 모형와 비교 대상모형의 평균제곱오차의 차이가 없다'이며 대립가설(H_1)은 'AFT 모형의 평균제곱오차가 비교 대상모형의 평균제곱오차 보다 작다'가 된다. 검정결과 유의

수준 0.05 미만의 p-value로 귀무가설을 기각하여, AFT 모형의 예측 오차가 평균 모형, 선형회귀 모형의 오차보다 작음을 확인할 수 있다. AFT 모형의 MSE 감소율은 <Table 6>과 같다.

Table 6. Performance comparison(MSE)

	AFT 모형 MSE	선형 모형 MSE	평균 모형 MSE	선형 모형대비 감소율	평균 모형대비 감소율
경로 1	30.93	32.98	54.62	6.2%	43.4%
경로 2	15.01	16.04	24.24	6.4%	38.1%
경로 3	22.41	23.16	28.66	3.2%	21.8%

모형의 제곱오차를 <Figure 6>과 같이 박스플롯(Box-plot) 그래프로 나타내면, AFT 모형이 선형회귀 모형과 평균 모형에 비해 오차의 중앙값(median)과 IQR(Inter-Quartile Range)이 작음을 직관적으로 알 수 있으며, <Table 7>과 같이 t검정으로 모형간의 유의차가 있음을 확인하였다.

Table 7. p-values of t-test on AFT with other models

	선형 모형	평균 모형
경로 1	< 0.001	< 0.001
경로 2	0.001	< 0.001
경로 3	0.002	< 0.001

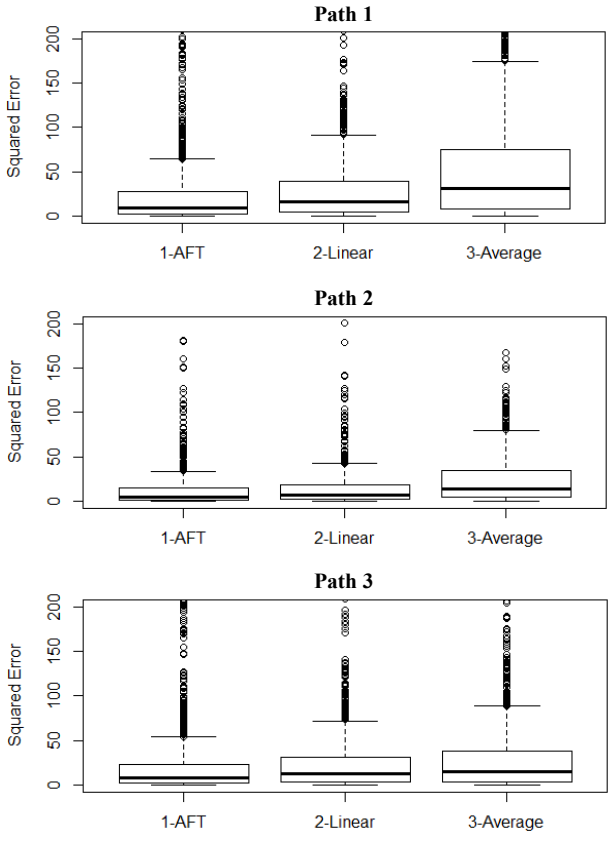


Figure 6. Comparison of model distribution

앞서 제 2장의 <Figure 4>에서 보았듯이 반송시간의 분포는 로그정규분포를 따르기 때문에 우편향된(right skewed) 오른쪽 꼬리부분이 긴 모양의 그래프가 그려지고, 이 부분이 <Figure 6>에서 이상치(Outlier)가 되어 위로 길게 퍼지는 것을 알 수 있다.

5. 결 론

본 연구에서는 디스플레이 제조공장에서 카세트 반송시간을 생존분석 방법을 사용하여 예측하는 모형을 제안하였다. 분포 가정 없이 모형의 계수를 추정하는 Cox 비례위험 모형으로 중요변수를 선택하고, 4가지 주요 생존시간 분포 중에 AIC 값을 기준으로 반송시간과 가장 적합한 분포를 찾은 후 최종적으로 반송시간을 예측하는 AFT 모형을 구축하였다. 결과적으로 평균제곱오차(MSE)가 기술 통계량을 사용한 예측방법 대비 평균 33.8%, 선형회귀 모형 대비 평균 5.3% 감소되었다.

본 연구에서 제안한 예측 모형은 스케줄링과 디스패칭 시스템의 생산계획과 분배성능을 개선할 수 있을 뿐만 아니라, 예측시간 오차로 발생하는 반송지연을 줄임으로서 설비 비가동과 생산성도 개선할 수 있을 것이다. 그리고 반송시간에 영향을 주는 중요변수만 다를 뿐 병렬생산을 하고 공장 내 재공재고의 공간 제약이 있는 전자, 반도체, 전지 산업 등에 여러 분야에 확대 적용할 수 있다.

추후연구로는 물류설비마다 다른 설비적인 고유특성이나 시간이 흐르면서 변화하는 시계열 변수(time dependent covariates)를 포함한 예측모형을 개발하는 것이 필요하다고 판단된다. 또한 본 연구에서 제안한 모형을 물류제어 시스템(AMHS)의 최적경로탐색 알고리즘에 활용함으로서 이동비용의 계산 척도를 현재 사용하는 반송거리가 아닌 반송시간을 기반으로 하는 경로탐색 방법에 대한 연구가 가능할 것으로 기대된다.

참고문헌

Akaike, H. (1974), A new look at the statistical model identification, *Automatic Control, IEEE Transactions on*, **19**(6), 716-723.

Bradburn, M. J., Clark, T. G., Love, S. B., and Altman, D. G. (2003), Survival analysis Part III : multivariate data analysis-choosing a model and assessing its adequacy and fit, *British Journal of Cancer*, **89**(4), 605.

Chen, T. (2009), A fuzzy-neural knowledge-based system for job completion time prediction and internal due date assignment in a wafer fabrication plant, *International Journal of Systems Science*, **40**(8), 889-902.

Chung, B. H. and Kim, S. I. (2008), Simulation Model for the Performance Analysis of In-Line Stocker System in TFT-LCD Fabrication Line, *Journal of the Korean Institute of Industrial Engineers*, 1-10

Cox, D. R. (1972), Regression models and life-tables, *Journal of the Royal Statistical Society, Series B (Methodological)*, 187-220.

Chung, B. Y., Heo, Y. H., Park, M. H., and Jang, S. C. (2010), Classification of the Standby time of Equipment to improve the productivity in

TFT-LCD Fab, *Journal of the Korean Institute of Industrial Engineers*, 1-6.

DeJong, C. D. and Wu, S. P. (2002), Material handling: simulating the transport and scheduling of priority lots in semiconductor factories, In *Proceedings of the 34th conference on Winter simulation : exploring new frontiers* (1387-1391). Winter Simulation Conference.

Fox, J. (2002), Cox proportional-hazards regression for survival data. *Appendix to An R and S-PLUS Companion to Applied Regression. Available at : http://cran.r-project.org/doc/contrib/Fox-Companion /appendix-cox-regression.pdf*.

Gu, S. J. and Choi, B. K. (2008), Simulation Model of In-Line Stocker System in TFT-LCD Line, *Journal of the Korean Institute of Industrial Engineers*, 1-7.

Hosmer Jr, D. W., Lemeshow, S., and May, S. (2011), *Applied survival analysis : regression modeling of time to event data*, Wiley.com, **618**.

Jang, Y. J. and Choi, G. H. (2006), Introduction to automated material handling systems in LCD panel production lines, *In Automation Science and Engineering*, CASE IEEE International Conference on (223-229), IEEE.

Jenkins, S. P. (2005), Survival analysis, *Unpublished manuscript*, *Institute for Social and Economic Research*, *University of Essex*, *Colchester*, *UK*.

Kong, S. H. (2007), Two-step simulation method for automatic material handling system of semiconductor fab, *Robotics and Computer-Integrated Manufacturing*, **23**(4), 409-420.

Lee, S. H., Moon, B. S., and Park, B. J. (2009), The Bus Delay Time Prediction Using Markov Chain, *The Journal of The Korea Institute of Intelligent Transport Systems*, **8**(3), 1-10.

Liao, D. Y. and Wang, C. N. (2004), Neural-network-based delivery time estimates for prioritized 300-mm automatic material handling operations, *Semiconductor Manufacturing*, *IEEE Transactions on*, **17**(3), 324-332.

Lin, J. S. and Wei, L. J. (1992), Linear regression analysis for multivariate failure time observations, *Journal of the American Statistical Association*, **87**(420), 1091-1097.

McCulloch, C. E. (2003), Generalized linear mixed models. In *NSF-CBMS Regional Conference Series in Probability and Statistics* (1-84), Institute of Mathematical Statistics and the American Statistical Association.

Qi, J. (2009), Comparison of proportional hazards and accelerated failure time models, *Doctoral dissertation, University of Saskatchewan*.

Shabtay, D. (2010), Scheduling and due date assignment to minimize earliness, tardiness, holding, due date assignment and batch delivery costs, *International Journal of Production Economics*, **123**(1), 235-242.

Shiue, Y. R. (2009), Data-mining-based dynamic dispatching rule selection mechanism for shop floor control systems using a support vector machine approach, *International Journal of Production Research*, **47**(13), 3669-3690.

Smith, R. M. and Bain, L. J. (1975), An exponential power life-testing distribution, *Communications in Statistics-Theory and Methods*, **4**(5), 469-481.

Son, H. S. and Choi, B. G. (2008), Capacity-Filtering Algorithm based Release Planning Method for TFT-LCD Fab, *Journal of the Korean Institute of Industrial Engineers*, 817-824.

Wei, L. J. (1992), The accelerated failure time model : a useful alternative to the Cox regression model in survival analysis, *Statistics in medicine*, **11**(14/15), 1871-1879.