DET SAMFUNDSVIDENSKABELIGE FAKULTET
Økonomisk Institut
KØBENHAVNS UNIVERSITET

Bus Engine Replacement Model and Heterogeneity.

Goutham Jørgen Surendran and Jesper Hybel Pedersen

Supervisor: Bertel Scherning. Associate Professor, PhD.

Department of Economics.

University of Copenhagen.

Contents

1 Introduction **3**

2 Framework for discrete dynamic processes **3**

2.1 Rust model 6

2.2 Unobserved Heterogeneity 7

2.3 Bus engine replacement model 7

3 Solving for EV **8**

3.1 Chebyshev polynomials 9

3.2 Bellman iterations 10

3.3 Newton-Kantorovich algorithm 12

4 Comparing integration techniques **13**

5 Estimation **13**

6 Monte Carlo Studies **16**

7 Applying the Function **19**

7.1 Simulating the expecting monthly bus engines replacements 20

7.2 Calculating using Rust(1987) “bottom up” approach 21

8 Conclusion **24**

9 References **25**

10 Appendix A: Integration technique results **26**

11 Appendix B: Gauss Hermite quadrature **28**

1 Introduction

Dynamic Discrete Choice Models (DDCMs) describe the behavior of a forward-looking economic agent who chooses between several alternatives repeatedly over time. Estimation of the deep structural parameters of these models is a theoretically appealing and promising area in empirical economics. One important feature of DDCMs that was often assumed away in the literature due to computational difficulties is heterogeneity in unobserved state variables. Ability, productivity, health status, taste idiosyncrasies, and many other unobservables are, however, likely to be present and be persistent over time.

An important part of solving stochastic dynamic optimization problems are the calculation of expectations, hence the evaluation of integrals. Usually these integrals lack closed form solutions and therefore numerical approximation techniques must be used. Since the integration are carried out in the inner most loop of the overall algorithms, used for solving the problem, speed and precision are crucial.

In this paper we solve the bus engine replacement problem with three different integration techniques: 1) Monte Carlo integration with Halton draws, 2) Equiprobable points and 3) Gauss Hermite quadrature and compare the techniques on speed and precision. We then develop a simple dynamic discrete choice model with heterogeneity and use Monte Carlo studies to investigate the consequences of not taking this heterogeneity into account when estimating the parameters of the model.

In section 2 we introduce the general dynamic optimization problem and derive the bellman equation heuristically. Then we introduce the three assumption of the Rust framework, which allows us to setup the contraction mapping used when solving for the expected value function EV and the choice probabilities of a conditional logit model in a dynamic setting. We then expand the framework to allow for unobserved heterogeneity in the instaneous payoff function and introduce heterogeneity in the busengine replacement model. In order to estimate parameters of the model we need to solve for EV and the algorithms used for this purpose are presented in section 3. Since these algorithms depend on integration techniques we make a small study of different integration techniques in section 4 to see the impact on speed and precision for our specific application. In section 5 we describe the likelihood functions used when estimating the parameters of the models after which we introduce a data generating proces in section 6 and carry out 3 small scale Monte Carlo studies to illustrate the properties of the MLE-estimators. Finnaly we calculate the expected demand using different techniques for the estimates of the Monte Carlo studies before we conclude in section 8.

2 Framework for discrete dynamic processes

In the general single agent framework for discrete dynamic processes we consider an agent who repeatedly faces a choice between a discrete set of complete and mutually exclusive alternatives $d_t \in D(s_t)$ as determined by the state variable s_t. Specifically we consider the simplest possible case of repeated binary choice $d_t \in \{0, 1\}$. The agent is assumed to make an infinite sequence

of decisions $\{d_{t+\tau}\}_{\tau=0}^{\infty}$ maximizing the expected sum of discounted instantaneous payoffs:

$$\max_{\{d_{t+\tau}\}_{\tau=0}^{\infty}} \mathbb{E}_t \left[\sum_{\tau=0}^{\infty} \beta^{\tau} U(s_{t+\tau}, d_{t+\tau}) \right] \tag{1}$$

The payoff function $U(.,.)$ is assumed constant and known to the decision-maker. The proces $\{s_{t+\tau}, d_{t+\tau}\}_{\tau=0}^{\infty}$ is a controlled markov proces with the transition probabilities given by $p(s_{t+1+\tau}|s_{t+\tau}, d_{t+\tau})$. The future values of the state variable are unknown to the decision-maker while the transition distribution is known and interpreted as representing the agents beliefs of future states. The initial period is indexed as period t such that s_t is assumed given. Maximizing the expected utility conditional on s_t the decision maker gets the lifetime payoff expressed by the indirect payoff function:

$$V(s_t) := \max_{\{d_{t+\tau}\}_{\tau=0}^{\infty}} \mathbb{E}_t \left[\sum_{\tau=0}^{\infty} \beta^{\tau} U(s_{t+\tau}, d_{t+\tau}) \right] \tag{2}$$

Since the expectation is taken with respect to the information set characterizing the state of knowledge of the agent in period t the payoff of the initial period can be moved outside the expectation operator:

$$= \max_{\{d_{t+\tau}\}_{\tau=0}^{\infty}} \left\{ U(s_t, d_t) + \beta \mathbb{E}_t \left[\sum_{\tau=0}^{\infty} \beta^{\tau} U(s_{t+1+\tau}, d_{t+1+\tau}) \right] \right\} \tag{3}$$

To maximize this sum with respect to the sequence $\{d_{t+\tau}\}_{\tau=0}^{\infty}$ is equivalent to maximizing with respect to the first period decision d_t and the sequence of future decisions $\{d_{t+1+\tau}\}_{\tau=0}^{\infty}$. Because the instantaneous payoff of the present period do not depend on the future decisions the problem can be restated as:

$$= \max_{\{d_t, \{d_{t+\tau}\}_{\tau=0}^{\infty}\}} \left\{ U(s_t, d_t) + \beta \mathbb{E}_t \left[\sum_{\tau=0}^{\infty} \beta^{\tau} U(s_{t+1+\tau}, d_{t+1+\tau}) \right] \right\} \tag{4}$$

$$= \max_{d_t} \left\{ U(s_t, d_t) + \beta \max_{\{d_{t+1+\tau}\}} \mathbb{E}_t \left[\sum_{\tau=0}^{\infty} \beta^{\tau} U(s_{t+1+\tau}, d_{t+1+\tau}) \right] \right\} \tag{5}$$

Using the fact that the least information set dominates $\mathbb{E}_t[.] = \mathbb{E}_t[\mathbb{E}_{t+1}[.]]$ and assuming the interchangeability of the max-operator and the $\mathbb{E}_t[.]$ we arrive at the expression:

$$= \max_{d_t} \left\{ U(s_t, d_t) + \beta \mathbb{E}_t \left[\max_{\{d_{t+1+\tau}\}} \mathbb{E}_{t+1} \left[\sum_{\tau=0}^{\infty} \beta^{\tau} U(s_{t+1+\tau}, d_{t+1+\tau}) \right] \right] \right\} \tag{6}$$

Imposing stationarity and inserting (x_t, ϵ_t) for s_t we thereby have the bellman equation:

$$V(x_t, \epsilon_t) = \max_{d_t} \{ U(x_t, \epsilon_t, d_t) + \beta \mathbb{E}_t [V(x_{t+1}, \epsilon_{t+1})] \} \tag{7}$$

Where we assume that x_t is observed by both agent and reseacher whereas ϵ_t is unobserved by the researcher but known by the decision-maker.

Given that the decision-maker maximizes expected utility the bellman equation suggests that the behavoir d_t of the decision-maker today can be explained on the basis of the observed variables $\{x_{t+\tau}\}_{\tau=0}^{\infty}$ and the unobserved variables $\{\epsilon_{t+\tau}\}_{\tau=0}^{\infty}$ and in particular through the way in which these variables impinge on the agents immidiate payoff $U(x_t, \epsilon_t, d_t)$ and the expected

payoff discounted $\beta\mathbb{E}_t\left[V(x_{t+1},\epsilon_{t+1})\right]$ given that future choices are optimal. The problem is thus twofold: (a) First of all the researcher is unable to use the unobserved variables as entering explanans since they are ex hypotesi unobserved and (b) Just as the future states of x are unknown to the agent so be they to the researcher.

Considering the problem of explaining the behavoir of the agent, problem (a) is in a way more a part of the solution than it is a problem. The optimal decision rule $d_t = \delta(x_t, \epsilon_t)$ giving the argument d_t that solves the maximization problem:

$$\delta(x,\epsilon) := \arg\max_{d_t} \left\{U(x_t,\epsilon_t,d_t) + \beta\mathbb{E}_t\left[V(x_{t+1},\epsilon_{t+1})\right]\right\} \tag{8}$$

would be a deterministic function $\delta(x_t)$ without the insertion of the unobserved variable. Since no theory is realistically capable of perfectly predicting the behavior of human decision-makers some remedy is necessary. The unobserved variables serves precisely the purpose of opening the door to probabilistic explanations by assuming ϵ_t to be stochastic. And because it is unlikely that any survey could completely record all information that is relevant to the agent's decision-making process the existence of unobserved state variables is quite plausible (Rust 1994: Handbook of Econometrics Vol. IV).

With respect to the problem (b) we first adopt the notation z' for time $t+1$ variables and z'' for time $t+2$ variables and write the one period ahead equation:

$$V(x',\epsilon') = \max_{d'} \left\{U(x',\epsilon',d') + \beta\mathbb{E}_{t+1}\left[V(x'',\epsilon'')\right]\right\} \tag{9}$$

and take the expectation with respect to time t:

$$EV(x,\epsilon,d) := \mathbb{E}_t[V(x',\epsilon')] = \int_{x'}\int_{\epsilon'} V(x',\epsilon')p(dx'd\epsilon'|x,\epsilon,d) \tag{10}$$

noting that this expectation is a function $EV(.)$ of the variables (x,ϵ,d) conditioned on. This allows us to express the bellman equation as:

$$V(x,\epsilon) = \max_{d} \left\{U(x,\epsilon,d) + \beta EV(x,\epsilon,d)\right\} \tag{11}$$

hence removing the future state variables by conditioning. The agents indirect payoff $V(x,\epsilon)$ is a function of todays state x given future optimal choices and assuming that the decision maker likewise optimizes when choosing the action d today.

Or put in another way: Given that the unobserved variable ϵ is distributed in the population according to a density $p(\epsilon)$, the optimal decision rule $\delta(x,\epsilon)$ is from the point of view of the researcher a stochastic variable. The agents behavoir today can therefore be probabilistically explained by noting that the probability, that the agent chooses a certain alternative $d \in D(x)$, must be equal to $Pr\big(\delta(x,\epsilon) = d|x\big)$, which can be found as the expectation to the indicator:

$$Pr\big(\delta(x,\epsilon) = d|x\big) = \int_{\epsilon} I\big\{d = \arg\max_{h}\left[U(x,\epsilon,h) + \beta EV(x,\epsilon,h)\right]\big\}p(d\epsilon) \tag{12}$$

taken with respect to the unobserved variables as they are distributed in the population. To make this model tractable further assumptions are called for.

2.1 Rust model

We now introduce three assumptions as presented by Rust(1987). The first assumption is the *conditional independence* limiting the pattern of dependence in the $\{x, \epsilon\}$ process in two ways:

$$p(x'\epsilon'|x, \epsilon, d) = p(\epsilon'|x'x, \epsilon, d)p(x'|x, \epsilon, d) = p(\epsilon'|x')p(x'|x, d) \tag{CI}$$

The reduction of $p(\epsilon'|x'x, \epsilon, d) = p(\epsilon'|x')$[1] tells us that x' is sufficient statistic for ϵ' ruling out serial dependence between ϵ' and ϵ. The reduction $p(x'|x, \epsilon, d) = p(x'|x, d)$ states that x' depends on x and not on ϵ. Which shows that the ϵ in $EV(x, \epsilon, d)$ is redundant such that we do not have to integrate $EV(.)$ with respect to ϵ. As a result taking the expectation of equation 9 and applying stationarity we have:

$$EV(x, d) := \int_{x'} \int_{\epsilon'} V(x', \epsilon')p(d\epsilon'|x')p(dx'|x, d) \tag{13}$$

$$= \int_{x'} \int_{\epsilon'} \max_{d'} \{U(x', \epsilon', d') + \beta EV(x', d')\}p(d\epsilon'|x')p(dx'|x, d) \tag{14}$$

The second assumption is that the unobserved variable enters the instantaneous payoff function *additive separably*:

$$U(x, \epsilon, d) = u(x, d) + \epsilon(d) \tag{15}$$

And the third assumption is that the unobserved variables $\{\epsilon_t\}_{t=0}^{\infty}$ are *independent and Extreme Value I distributed* conditional x such that the integral over ϵ' has the closed form solution known as the social value function. This means that the expectation of future indirect payoffs can be expressed as the integral to the closed form solution with respect to the transition density:

$$EV(x, d) = \Gamma(EV)(x, d) = \int_{x'} \log \sum_{d'} \exp\{u(x', d') + \beta EV(x', d')\}p(dx'|x, d) \tag{16}$$

The inner expression $u(x', d') + \beta EV(x', d')$ depends only on d' and x'. The choice variable d' is removed by summing over the alternatives, in this case $d' = 0$ for keep or $d' = 1$ for replacement. And the state variable milage x' is integrated out. As a result the integral only depends on x and d of the conditional distribution. Furthermore it can be shown that the expression is a contraction mapping with $EV(.)$ as fixpoint.

The probability of the agent choosing a specific alternative d is given as:

$$Pr\big(\delta(x, \epsilon) = d|x\big) = \int_{\epsilon} I\big\{d = \arg\max_{h} \big[u(x, h) + \epsilon(h) + \beta EV(x.h)\big]\big\}p(d\epsilon|x) \tag{17}$$

and by applying the result of McFadden(1981) using that $\{\epsilon_t|x_t\}_{t=0}^{\infty}$ is a independent Extreme Value I process the choice probabilities can be expressed as:

$$Pr(d|x) = \frac{\exp\{u(x, d) + \beta EV(x, d)\}}{\sum_j \exp\{u(x, j) + \beta EV(x, j)\}} = \frac{\exp\{v(x, d)\}}{\sum_j \exp\{v(x, j)\}} \tag{18}$$

defining alternative specific value function $v(x, j) := u(x, j) + \beta EV(x, j)$. With $\beta = 0$ the agent disregards future payoffs and the model becomes equivalent with a static logit model.

[1] Usually it is further assumed that $p(\epsilon'|x') = p(\epsilon')$.

2.2 Unobserved Heterogeneity

The Rust model can be extended in order to take unobserved heterogeneity into account by assuming that the payoff function depends on an unobserved component ξ specific to the unit of observation n:

$$U(x, \epsilon, \xi_n, d) = u(x, \xi_n, d) + \epsilon(d) \tag{19}$$

The distributional assumption for the model can be stated as:

$$p(x'\epsilon'|x, \epsilon, \xi_n, d) = p(\epsilon'|x', \xi_n)p(x'|x, d) \tag{20}$$

as a consequence we get an unit specific expected value function EV using:

$$EV(x, d, \xi_n) := \int_{x'} \int_{\epsilon'} V(x', \epsilon', \xi_n)p(\epsilon'|x', \xi_n)p(x'|x, d) \tag{21}$$

For a given value of the unit specific variable the problem of solving for the expected value function reduces to the problem of solving for $EV(x, d)$ in the Rust model. The expected value function can therefore be found as the solution to the contraction mapping:

$$EV(x, d, \xi_n) = \Gamma(EV)(x, d, \xi_n) \tag{22}$$

$$= \int_{x'} \log \sum_{d'} \exp\{u(x', d', \xi_n) + \beta EV(x', d', \xi_n)\}p(dx'|x, d) \tag{23}$$

While the simulation of data requires solving for the unit specific expected value function in order to calculate the probabilies:

$$Pr(d_{nt}|x_{nt}, \xi_n) = \frac{\exp\{u(x_{nt}, d_{nt}, \xi_n) + \beta EV(x_{nt}, d_{nt}, \xi_n)\}}{\sum_j \exp\{u(x_{nt}, j, \xi_n) + \beta EV(x_{nt}, j, \xi_n)\}} \tag{24}$$

the estimation must be based on observables only:

$$Pr(\mathbf{d}_n|\mathbf{x}_n) = \int_{\xi} \prod_{t=1}^{T_t} Pr(d_{nt}|x_{nt}, \xi)p(d\xi|x_{nt}) \tag{25}$$

hence we integrate out the unobserved heterogeneity when using the partial likelihood $l^3(.)$ as set up in section 5.

2.3 Bus engine replacement model

In the bus engine replacement model the agent oversees a fleet of buses $n = 1, ..., N$, each period observing the accumulated milage x_{tn} of bus n for period t. The state variable of accumulated milage x is assumed to be a regenerative random walk with the change $\eta = x' - x$ following a density $g(.; \theta_1)$ depending on the parameters θ_1:

$$p(x'|x, d) = \begin{cases} g(x' - x; \theta_1) & \text{if } d = 0 \\ g(x'; \theta_1) & \text{if } d = 1 \end{cases} \tag{26}$$

The accumulated bus milage cannot but decrease, nevertheless we assume the density to be normal:

$$g(.; \theta_1) = \mathcal{N}(\mu_1, \sigma_1) \tag{27}$$

We use this choice of density for several reasons: 1) it is well known and hence particularly simple to work with, 2) the data appears approximately normal, 3) it is computationally non-demanding to integrate out using Gauss-Hermite quadrature and 4) it is easy to quasi-bound ensuring values above 0 and smaller than some constant K simly by choosing a given mean μ_1 and a compared to the mean realtively small σ_1. Using a lognormal distribution would be an option ensuring positive support, but a drawback of this density for the particular problem at hand are the heavy tails, making it hard to ensure that the accumulated milage do not exceed the grid approximating the state space for the upper bound of the grid.

In each period the agent after having observed the accumulated milage decides whether to replace the engine of the bus $d_{nt} = 1$ or instead do regular maintenance $d_{nt} = 0$. The resulting payoff is captured by the function $u(.)$ which we assume to be linear:

$$u(x,d) = \begin{cases} -0.001 \cdot \theta_c x & \text{if } d = 0 \\ -RC & \text{if } d = 1 \end{cases} \tag{28}$$

and increasing in milage x with respect to the maintanence costs $\theta_c \cdot x$ reflecting the effect of wear and tear. In the models with heterogeneity we assume that busdrivers, assigned to the same busses, displays differing levels of recklessness in their driving. We model this by assuming that the cost coefficient is normally distributed:

$$\theta_{nc} = \mu_c + \sigma_c z_n \qquad z_n \sim \mathcal{N}(0,1) \tag{29}$$

The mean and standard deviation are constant across time and busses hence characterizing the population and the coefficients are only varying over busses or equivalently over busdrivers.

3 Solving for EV

In this section we first present the method we have used for making a continuous differentiable approximation of EV using chebyshev polynomials. The method presented here is more thoroughly described in Judd (1998: 238) We then explain how we use this approximation when implementing bellman iterations and Newton Kantorovich iterations. Together these two algorithms constitutes the polyalgorithm presented by John Rust (2000) in the NFXP manual.

While the contraction iterations are guaranteed to converge the method is slow especially with high values of β. On the other hand convergence of the NK-algorithm depends on being in an area of attraction in which case the algorithm is relative fast, converging in a couple of iterations. The trick is to get as few value function iterations as possible by changing to NK-iteration as soon as the algorithm enters the area of attraction, usually occuring after 20-30 iterations.

Needless to say the following section depends heavily on the NFXP manual by John Rust (2000) in which he also comments further on the aspect of flow control in the application of the algorithm. We also rest heavily on the understanding gotten from reading the code by Bertel Scherning applying the NFXP to discretized data as in the NFXP manual. We adapt these resources for continuous data and unlike Rust we choose to use a numerical approximation of the Frechet derivative of the contraction mapping because we found this method to be easier.

3.1 Chebyshev polynomials

In order to solve the model $\mathcal{M}(\boldsymbol{\theta})$ for EV_θ taking the parameters $\boldsymbol{\theta} = (RC, \theta_c, \mu_1, \sigma_1)$ and β for granted we first set up a grid $\check{x} := (\check{x}_1, ..., \check{x}_G)$ for the milage state variable x. Since we use chebyshev approximation for EV_θ on the grid this is done by choosing G the number of gridpoints along with upper and lower bounds $[l, u]$ of which the gridpoints represents a discretization. The chebyshev nodes are then given as:

$$\check{c}_g = -\cos\left(\frac{\pi(2g-1)}{2G}\right) \qquad g = 1, ..., G \tag{30}$$

and are located on the interval $[-1, 1]$. These notes are then transformed to get $\check{x}$ using the linear transformation:

$$\check{x}_g = l + \frac{(\check{c}_g + 1) * (u - l)}{2} \tag{31}$$

We then choose polynomial order $Q - 1$ of the chebyshev approximation and for each node $\check{c}_g$ we then find the $1 \times Q$ chebyshev vector $T_g = (T_{1g}, T_{2g}, ..., T_{Qg})^\top$ using the recursion over coordinates:

$$T_{qg} = 2\check{c}_g T_{q-1,g} - T_{q-2,g} \qquad q = 1, ..., Q - 1 \tag{32}$$

with $T_{0g} = 1$ and $T_{1g} = \check{c}_g$. Alternatively one can use the formula:

$$T_{qg} = \cos((q-1) * \text{acos}(\check{c}_g))$$

Having found the vectors $T_1, ..., T_G$ they are stabled to give the $G \times Q$ matrix T and calculate the $Q \times G$ matrix H:

$$T := \begin{bmatrix} T_1^\top \\ | \\ T_G^\top \end{bmatrix} \qquad H := (T^\top T)^{-1} T^\top \tag{33}$$

We then define $T_x^\top$ as the $1 \times Q$ vector resulting from first applying the inverse of 31 to x giving us c:

$$c(x) = \frac{2(x - l)}{(u - l)} - 1 = c \tag{34}$$

and then applying the recursion 32 to c by substituting $\check{c}_g$ with c in 32. Finally we can define $\widetilde{EV}_\theta^b(x)$ using the interpolating transformation $\widetilde{EV}_\theta^b(x) : x \in [l, u] \to \mathbb{R}$ given as:

$$\widetilde{EV}_\theta^b(x) = T_x^\top (T^\top T)^{-1} T^\top \check{EV}_\theta^{\,b} \tag{35}$$

with $\check{EV}_\theta^{\,b} := (\check{EV}_1^b, ..., \check{EV}_G^b)^\top$ being the vector for the expected value function on the grid $\check{x}$ gotten from applying the bellman iteration b times [2]. When $b = 0$ the vector $\check{EV}_\theta^{\,0}$ is simply our initial guess for EV_θ and with the interpolation at hand this guess is not only a guess for EV_θ on the grid but a guess for EV_θ in any point x on the interpolation interval $[l, u]$.

[2] More precisely we have one vector for each alternative $d \in D$ but as argued later we only need to consider one alternative.

While it is unecessary - from a mathematical perspective - to apply $c(x)$ to x in order to transform x from the interval $[l, u]$ to $[-1, 1]$ it is necessary from a computational point of view in order to prevent numerical overflow. It could also be noted that since the chebyshev polynomials are orthogonal $T^\top T$ is diagonal $T^\top T = \text{diag}(t_1, ..., t_Q)$ such that $(T^\top T)^{-1} = \text{diag}(1/t_1, ..., 1/t_Q)$ and furthermore $t_1 = G$ and $t_{q\neq 1} = G/2$.

- Given l, u, G and $\check{EV}_\theta^{\,b}$ chebyshev approximation gives us $\widetilde{EV}_\theta^b(x) : x \in [l, u] \to \mathbb{R}$. While T_x depends on interpolated x and $\check{EV}_\theta^{\,b}$ depends on bellman iteration b the matrix $(T^\top T)^{-1}T^\top$ is independent of either and can therefore be calculated before the bellman iteration.

3.2 Bellman iterations

Having set up the grid by choosing l, u, G and chosen an inital guess for $\check{EV}_\theta^{\,0}$ we now have:

$$\forall x \in [l, u] : \quad \widetilde{EV}_\theta^b(x) = T_x^\top (T^\top T)^{-1} T^\top \check{EV}_\theta^{\,b} \tag{36}$$

as a function of x for $b = 0$. We then use the contraction mapping:

$$EV(x, d) = \Gamma(EV)(x, d) = \int_{x'} \log \sum_{d'} \exp\{u(x', d') + \beta EV(x', d')\} p(dx'|x, d) \tag{37}$$

In order to make this applicable we use the property of the regenerative random walk implying that:

$$\forall x : \qquad EV(x, d = 1) = EV(0, d = 0) \tag{38}$$

This means that we can define the function $EV(x) := EV(x, 0)$ and only need to consider the solution of $EV(x)$. We only therefore only need one vector $\check{EV}_\theta^{\,b}$ in each bellman iteration approximating the function $EV_\theta(x, d)$. This allows us to rewrite the contraction mapping be seeting $d = 0$ and summing over d':

$$EV(x, 0) = \int_{x'} \log\left(\exp\{u(x', 0) + \beta EV(x', 0)\} + \exp\{u(x', 1) + \beta EV(x', 1)\}\right) p(dx'|x, 0) \Leftrightarrow$$

$$EV(x) = \int_{x'} \log\left(\exp\{u(x', 0) + \beta EV(x')\} + \exp\{u(x', 1) + \beta EV(0)\}\right) p(dx'|x, 0) \Leftrightarrow$$

We the apply the definition of the payoff function for the bus engine replacement model in order to get:

$$EV(x) = \int_{x'} \log\left(\exp\{-\theta_c x' + \beta EV(x')\} + \exp\{-RC + \beta EV(0)\}\right) p(dx'|x, 0) \tag{39}$$

and change the variable of integration using that $x' = x + \eta$:

$$EV(x) = \int_{\eta} \log\left(\exp\{-\theta_c(x + \eta) + \beta EV(x + \eta)\} + \exp\{-RC + \beta EV(0)\}\right) g(d\eta) \tag{40}$$

since we have chosen η to be a normally distributed variable the function we need to integrate is a function of a normally distributed stochastic variable and hence numerically integrable by Gauss Hermite quadrature.

Numerically we have two dimensions: 1) The dimension of the grid $\check{x} = (\check{x}_1, ..., \check{x}_G)$ and 2) The dimension of the Gauss Hermite nodes $\mathbf{v} = (v_1, ..., v_H)$. So we set up the matrix:

$$\mathbf{V} = \begin{bmatrix} \check{x}_1 + v_1 & ... & \check{x}_1 + v_H \\ \check{x}_2 + v_1 & ... & \check{x}_2 + v_H \\ \vdots & \ddots & \vdots \\ \check{x}_G + v_1 & ... & \check{x}_G + v_H \end{bmatrix} \tag{41}$$

Define the vector of weights $\mathbf{w} = (w_1, ..., w_H)^\top$ and the transformation $F(.) : G \times H \to G \times H$ that applies the function $G(y)$ to each element of the matrix $\mathbf{V}$:

$$F(\mathbf{V}) = \begin{bmatrix} G(\check{x}_1 + v_1) & ... & G(\check{x}_1 + v_H) \\ G(\check{x}_2 + v_1) & ... & G(\check{x}_2 + v_H) \\ \vdots & \ddots & \vdots \\ G(\check{x}_G + v_1) & ... & G(\check{x}_G + v_H) \end{bmatrix} \tag{42}$$

With $G(y)$ short of recentering being defined as[3]:

$$G(y) = \log\Big(\exp\{-\theta_c y + \beta\widetilde{EV}_\theta^b(y)\} + \exp\{-RC + \beta\widetilde{EV}_\theta^b(0)\}\Big) \tag{43}$$

$$= \log\Big(\exp\{-\theta_c y + \beta T_y^\top (T^\top T)^{-1} T^\top \check{EV}_\theta{}^b\} + \exp\{-RC + \beta T_0^\top (T^\top T)^{-1} T^\top \check{EV}_\theta{}^b\}\Big)$$

We therefore have, as a numerical version of the contraction mapping given in equation 40 the matrix calculation:

$$\check{EV}_\theta^{b+1} = \frac{1}{\sqrt{\pi}} F(\mathbf{V})\mathbf{v} \tag{44}$$

which implements one bellman iteration.

To recapitulate we have pressuposed the inputs given in table 1. And we use T_y which is constant because the y's are elements of the matrix $\mathbf{V}$ being constant. We have used $(T^\top T)^{-1}T^\top$ which is constant. And we have used $\theta_c \mathbf{V}$ inside $G(y)$ also being constant.

Table 1: Inputs for bellman iteration

Input	
$\boldsymbol{\theta} = (RC, \theta_c, \mu_1, \sigma_1)$	Model parameters
l	lower bound on grid
u	upper bound of grid
G	the number of gridpoints
$\check{EV}_\theta{}^0$	Initial guess for $EV_\theta(x)$
$\mathbf{v}$	Gauss Hermite nodes
$\mathbf{w}$	Gauss Hermite weights

While the bellman iterations will converge the convergence can be slow as already explained, hence we also use Newton Kantorovich iterations as described in the next section (Rust 2000: 20).

[3]Recentering is necessary in order to avoid numerical under- or overflow. See NFXP manual equation 3.20 (Rust 2000: 27).

3.3 Newton-Kantorovich algorithm

The solution of EV_θ is a fixed point of the contraction mapping $EV_\theta = \Gamma(EV_\theta)$ which holds if and only if $[I - \Gamma]EV_\theta = 0$. The operator $F = [I - \Gamma]$ has Frechet derivative $F' = [I - \Gamma']$ with the inverse $[I - \Gamma']^{-1}$. By first order Taylor approximation we have

$$0 = F(EV_\theta^{b+1}) \approx F(EV_\theta^b) + F'(EV_\theta^b)(EV_\theta^{b+1} - EV_\theta^b) \tag{45}$$

Solving for EV_θ^{b+1} we get:

$$0 \approx F(EV_\theta^b) + F'(EV_\theta^b)(EV_\theta^{b+1} - EV_\theta^b) \Leftrightarrow \tag{46}$$

$$-F'(EV_\theta^b)(EV_\theta^{b+1} - EV_\theta^b) \approx F(EV_\theta^b) \Leftrightarrow \tag{47}$$

$$EV_\theta^{b+1} \approx EV_\theta^b - F'^{-1}(EV_\theta^b)F(EV_\theta^b) \tag{48}$$

We then use that $F(EV_\theta^b) = EV_\theta^{b+1} - EV_\theta^b$ and that $F'^{-1}(EV_\theta^b) = [I - \Gamma']^{-1}$ with Γ' evaluated for EV_θ^b:

$$EV_\theta^{b+1} \approx EV_\theta^b - [I - \Gamma'(EV_\theta^b)]^{-1}(EV_\theta^{b+1} - EV_\theta^b) \tag{49}$$

This suggest the following procedure:

1. Given $\check{EV}_\theta^b$ found using bellman iterations calculate $\Gamma'(\check{EV}_\theta^b)$.
2. Calculate $\check{EV}_\theta^{b+1}$ using bellman iteration on $\check{EV}_\theta^b$.
3. Take one NK step by: Inserting $\check{EV}_\theta^b$, $\check{EV}_\theta^{b+1}$ and $\Gamma'(\check{EV}_\theta^b)$ on the right hand side of equation 49 calculate left hand side $\check{EV}_\theta^{b+1}$ of the equation.
4. Take one bellman step based on the previously calculated LHS $\check{EV}_\theta^{b+1}$ to find $\check{EV}_\theta^{b+2}$ and test for acceptance. If acceptance go to step 1 with the LHS $\check{EV}_\theta^{b+1}$ used in the role of $\check{EV}_\theta^b$ and with $\check{EV}_\theta^{b+2}$ in the role of $\check{EV}_\theta^{b+1}$. If non-acceptance return to bellman doing iterations until reentering area of attraction.

Since $\check{EV}_\theta^b$ is a vector we approximate $\Gamma'(\check{EV}_\theta^b)$ by approximating the Jacobi matrix of Γ. This is done by using the newton fraction, where we approximate column k of the Jacobi using:

$$J_k(\Gamma(\check{EV}_\theta^b)) \approx \frac{\Gamma(\check{EV}_\theta^b + h \cdot e_k) - \Gamma(\check{EV}_\theta^b)}{h} \tag{50}$$

with h being a small scalar - we found $h = 0.000001$ to work nicely - and e_k is the euclidean $G \times 1$ basis vector with k'th coordinate equal to one. The cookbook procedure is as follows:

1. Given $\check{EV}_\theta^b = (\check{EV}_1^b, ..., \check{EV}_G^b)$ construct $\check{EV}_\theta^b + h \cdot e_k$.
2. Apply bellman iteration to $\check{EV}_\theta^b + h \cdot e_k$ to get $\Gamma(\check{EV}_\theta^b + h \cdot e_k)$.
3. Using $\check{EV}_\theta^{b+1} = \Gamma(\check{EV}_\theta^b)$ calculate $\Gamma(\check{EV}_\theta^b + h \cdot e_k) - \check{EV}_\theta^{b+1}$ and divide with h.
4. Repeat step 1-3 for $k = 1, ..., G$ getting $\check{J}_1, ..., \check{J}_G$
5. Let the approximation of $\Gamma' \approx [\check{J}_1, ..., \check{J}_G]$ being a $G \times G$ matrix.

In order to apply the algorithm a criteria of when the bellman iterations is possible in an area of attraction has to be decided. This criteria determines when the algorithm changes from doing bellman iterations and starts doing NK-iterations. And as mentioned a criteria of acceptance of the NK-steps has to be decided[4].

4 Comparing integration techniques

One of the main advantages of using chebyshev approximation for EV_θ and solving the model with a continuous state variable is that it becomes possibly to use Monte Carlo integration. While this integration technique is very general, gains in precision are costly in terms of the number of draws and hence computation time. Heuristically this can be understood as being an effect of having an inefficient spread of draws on the support. From sample selection theory it is known that the simple sample average can be improved by taking relatively larger subsamples from high variance subregions of the support. The Monte Carlo integration technique »looses« precision due to the oversampling of high probability low variance areas. One way to prevent this oversampling is to introduce negative correlation between the draws as is done with Halton draws. Another way is to select specific - rather than randomly sample - points from the support and then weigh these with the proper weights. The method of equiprobable points as well as Gauss-Hermite quadrature are both examples of this approach, with the first being more general than the latter.

To evaluate the different integration techniques we first solve the model for EV using the relevant integration technique and calculate the choice probabilies $Pr(d|x)$ for a vector of x-values $(x_1, ..., x_S)$. These choice probabilites are compared with the choice probabilities $Pr_0(d|x)$ found using EV_0, where EV_0 is the solution of the expected value function found using Gauss Hermite integration with 100 Gauss Hermite nodes and 50 chebyshev nodes for a polynomial approximation of order 49. The result is two vectors of choice probabilities $\mathbf{p} = (Pr(d|x_1), ..., Pr(d|x_S))$ and $\mathbf{p}_0 = (Pr_0(d|x_1), ..., Pr_0(d|x_S))$ on the basis of which we find the $error := 100 \cdot \max|\mathbf{p} - \mathbf{p}_0|$ as the max deviance in percentage points.

To illustrate the trade off between speed and precision we also time the calculation of EV for each integration technique and calculate the measure:

$$E = \frac{\exp(-\text{error})}{\exp(-\text{error}) + \exp(\text{time})} \in (0, 1/2) \tag{51}$$

which is 0.5 for a very fast and precise technique and 0 for a slow and imprecise technique. The results are illustrated in table 2, 3 and 4, indicating that a use of 10 chebyshev nodes polynomial order of 9 are optimal. The trade off gets worse for higher number of draws or notes because the gain in precision is very small compared to the extra time used for model solution [5].

5 Estimation

For estimation of the parameters the partial loglikelihood functions are used in a two step procedure.

[4]For criteria of acceptance see NFXP manual comments on flow control pp. 28-30

[5]See tables in appendix A showing only the error.

Table 2: *Gauss Hermite Integration*

	Number of Gauss Hermite nodes						
C-Nodes	2	4	8	16	32	64	100
2	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4	0.04	0.04	0.04	0.04	0.04	0.04	0.04
6	0.41	0.41	0.41	0.41	0.41	0.40	0.40
8	0.30	0.30	0.29	0.29	0.29	0.29	0.29
10	0.48	0.48	0.48	0.48	0.47	0.46	0.46
20	0.48	0.48	0.48	0.47	0.46	0.44	0.41
50	0.47	0.46	0.45	0.41	0.33	0.16	0.09

Note 1: c-nodes are the number of chebyshev nodes

Note 2: Transitionparameters $\mu_1 = 7$ and $\sigma_1 = 1$

Table 3: *Equiprobable Integration*

	Number of nodes					
C-nodes	10	50	100	500	1000	5000
2	0.00	0.00	0.00	0.00	0.00	0.00
4	0.04	0.04	0.04	0.03	0.02	0.00
6	0.41	0.40	0.39	0.33	0.26	0.03
8	0.29	0.29	0.28	0.23	0.17	0.01
10	0.47	0.46	0.45	0.36	0.29	0.02
20	0.47	0.45	0.42	0.24	0.11	0.00
50	0.44	0.31	0.16	0.00	0.00	0.00

Note 1: c-nodes are the number of chebyshev nodes

Note 2: Transitionparameters $\mu_1 = 7$ and $\sigma_1 = 1$

Table 4: *Monte Carlo integration with Halton draws*

	Number of draws					
C-nodes	10	50	100	500	1000	5000
2	0.00	0.00	0.00	0.00	0.00	0.00
4	0.03	0.04	0.03	0.04	0.03	0.03
6	0.40	0.41	0.40	0.38	0.36	0.22
8	0.28	0.28	0.27	0.27	0.24	0.12
10	0.47	0.46	0.45	0.41	0.38	0.14
20	0.46	0.45	0.43	0.29	0.16	0.00
50	0.42	0.31	0.16	0.00	0.00	0.00

Note 1: c-nodes are the number of chebyshev nodes

Note 2: Transitionparameters $\mu_1 = 7$ and $\sigma_1 = 1$

First the loglikelihood $l^1(.) = \sum_{n=1}^{N} l_n^1(.)$ is maximized to estimate the parameters $\boldsymbol{\theta_1} = (\mu_1, \sigma_1)$ of the density $p(x'|x, d; \boldsymbol{\theta}_1)$ governing the regenerative random walk of bus milage x.

Secondly the loglikelihood $l^2(.) = \sum_{n=1}^{N} l_n^2(.)$ is used for estimation of $\boldsymbol{\theta}_2 = (RC, \theta_c)$. When estimation is carried out taking into account the heterogeneity of the cost function coefficient $\theta_{nc} = \mu_c + \sigma_c z_n$ we use $l^3(.) = \sum_{n=1}^{N} l_n^3(.)$ in the second step for estimating $\boldsymbol{\theta}_3 = (RC, \mu_c, \sigma_c)$.

While the partial loglikelihoods can be used for consistent estimation of parameters, they cannot be used for calculating a consistent estimator of the asymptotic covariance matrix of the estimator (Rust 1978: 1019). However Rust notes that he found the estimated covariance matrix of l^2 to be almost perfectly block diagonal. This suggest that for all practical purposes it is

unnecessary to use l^f although it produces efficient maximum likelihood estimates and estimates the covariance matrix consistently. Nevertheless we use the procedure of finding initial estimates based on the partial likelihoods l^1, l^2 - or l^3 in the case of heterogeneity - after which we initialize the full maximum likelihood estimator with these estimates.

The data is assumed to consist of N independent observation of buses such that the likelihood function $L_N(.)$ is given as the product of $n = 1, ..., N$ individual likelihood contributions. For the full maximum likelihood the individual likelihood contribution for bus n can be written as:

$$L_n(\mathbf{x}_n, \mathbf{d}_n | x_0, d_0; \boldsymbol{\theta}) = \prod_{t=1}^{T_n} Pr(d_t | x_t; \boldsymbol{\theta}) g(x_t | x_{t-1}, d_{t-1}; \boldsymbol{\theta}) \tag{52}$$

with $\mathbf{d}_n = (d_1, ..., d_{T_n})$ and $\mathbf{x}_n = (x_1, ..., x_{T_n})$. The full loglikelihood $l(.) = \log L_n(.)$ can be written as the sum of the two partial log likelihoods $l_n^1(.)$ and $l_n^2(.)$ given as:

$$l_n^1(\mathbf{d}_n, \mathbf{x}_n; \boldsymbol{\theta}_1) = \sum_{t=1}^{T_n} \log g(x_t | x_{t-1}, d_{t-1}; \boldsymbol{\theta}_1) \tag{53}$$

$$l_n^2(\mathbf{d}_n, \mathbf{x}_n; \boldsymbol{\theta}_2) = \sum_{t=1}^{T_n} \log Pr(d_{nt} | x_{nt}; \boldsymbol{\theta}_2) \tag{54}$$

Without heterogeneity the parameters are $\boldsymbol{\theta} := (RC, \theta_c, \mu_1, \sigma_1)$ excluding β since we do not estimate this parameter. We choose to write the choice probabilities as depending only on $\boldsymbol{\theta}_2 = (RC, \theta_c)$. Implicitly however they do depend on $\boldsymbol{\theta}_1 = (\mu_1, \sigma_1)$ through $EV_\theta(x, d)$. The estimation procedure is therefore to estimate $\boldsymbol{\theta}_1$ first using $l_n^1(.)$. Using $\hat{\boldsymbol{\theta}}_\mathbf{1}$ it is possible to solve for $EV_{\hat{\theta}}(x, d)$ for given values of (RC, θ_c) and β implying that a new estimate of $EV_{\hat{\theta}}(x, d)$ is solved for in each evaluation of $l_n^2(.)$ carried out by the solver routine.

Adding heterogeneity to the cost coefficient $\theta_{nc} = \mu_c + \sigma_c z_n$ with $z_n \sim \mathcal{N}(0, 1)$ leaves the markov process of the state variable x unchanged. This means that $\boldsymbol{\theta}_1$ still can be consistently estimated using $l_n^1(.)$. The choice probabilities conditional on the state variable can be found integrating out the unobserved heterogeneity:

$$Pr(\mathbf{d}_n | \mathbf{x}_n; \boldsymbol{\theta}_3) = \int_{\theta_c} \prod_{t=1}^{T_n} Pr(d_{nt} | x_{nt}; \theta_c, \boldsymbol{\theta}) f(\theta_c) d\theta_c \tag{55}$$

$$= \int_z \prod_{t=1}^{T_n} Pr(d_{nt} | x_{nt}, z; \boldsymbol{\theta}_3) \phi(z) dz \tag{56}$$

Substituting this expression $Pr(\mathbf{d}_n | \mathbf{x}_n; \boldsymbol{\theta}_3)$ for $Pr(\mathbf{d}_n | \mathbf{x}_n; \boldsymbol{\theta})$ in the expression for the full likelihood function and taking logs we have $l_n(.) = l_n^3(.) + l_n^1(.)$ where:

$$l_n^3(\mathbf{d}_n, \mathbf{x}_n; \boldsymbol{\theta}_3) = \log \int_z \prod_{t=1}^{T_n} Pr(d_{nt} | x_{nt}, z; \boldsymbol{\theta}_3) \phi(z) dz \tag{57}$$

and $\boldsymbol{\theta}_3 = (RC, \mu_c, \sigma_c)$ now excluding θ_c. The integral can be approximated using Monte Carlo simulation but we choose to use Gauss-Hermite quadrature since this probably results in an algorithm using fewer solutions for EV_θ. The reason for this is that with heterogeneity, the function $EV_\theta(.)$ is also a function of ξ. In the present model $\xi_n = \theta_{nc}$ the unit specific cost coefficient.

For each value of ξ it is necessary to solve for $EV_\theta(.)$ suggesting that using Monte Carlo integration with R draws would require R solutions of EV_θ per iteration of likelihood evaluations. It could be argued that EV_θ changes very little, for relatively small changes in θ_c, suggesting the procedure of randomly generating $(\theta_c^{(1)}, ..., \theta_c^{(R)})$ sorting and then only recalculating EV_θ for every k'th θ_c^r [6].

6 Monte Carlo Studies

To investigate the behavior of the different estimators we perform three Monte Carlo studies as listed in table 5. In the first study we generate data without heterogeneity and use the maxmimum likelihood estimation assuming no heterogeneity. In the second study we generate data with heterogeneity and investigates the consequences of estimating under the assumption that there is no heterogeneity. In the third and final study we generate data with heterogeneity and estimate the parameters of the model assuming heterogeneity.

To generate the data assuming heterogeneity we use the following data generating proces:

1. Make a draw of $\xi^n \sim \mathcal{N}(\mu_c, \sigma_c)$.

2. Insert ξ^n in $EV(.)$ and let $d = 0$ in order to get the function

$$EV(x, d = 0, \xi = \xi^n) : (x) \to \mathbb{R} \tag{58}$$

   using chebyshev polynomials this is done by making an initial guess for $\check{EV}^0$ and solving for $EV(x|d, \xi)$ using $EV(x, 0, \xi^n) = \Gamma(EV)(x, 0, \xi^n)$ as given in equation 23. This mean we apply the polyalgorithm of NFXP using contraction mappings and Newton Kantorovich iterations.

3. Having solved for $EV(x|d = 0, \xi^n)$ we can calculate the choice probabilities for a given value of x. Hence we initialize a proces $\{x_t^n\}_{t=0}^T$ in a chosen x_0 and find $x_1^n = x_0^n + \eta_1$ where η_1 is a draw $\eta \sim \mathcal{N}(\mu_1, \sigma_1)$. Calculate $Pr(dr_1|x_1^n, \xi^n)$ and let $d_1^n = I[u < Pr(d_1|x_1^n, \xi^n)]$ where $u \sim \mathcal{U}(0, 1)$. Assuming the calculated probability is the probability of replacement we change x_1 to equal 0 if the $d_1 = 1$ after which we calculate $x_2^n = x_1^n + \eta_2$ same procedure.

4. The proces is repeated until T period storing $d_1^n, ..., d_T^n$ and $x_0^n, x_1^n, ..., x_T^n$ for N replicates of units $n = 1, ...N$.

To genereate the data without heterogeneity the above algorithm is used with $\xi^r = \theta_c \ \forall r$ such that step 1 simply involves choosing a value for θ_c and step 2 is only performed once. Each data generating proces - DGP - is repeated B times.

The results of the three studies are given in table 6 and are commented in the following three sections.

MC-study I

In the first Monte Carlo study we simulate $B = 2500$ data sets without heterogeneity and estimate RC and θ_c using the full maximum likelihood estimates. Basd on the sequence of

[6] For a comment on this see Scherning (2007) footnote 11.

Table 5: *Monte Carlo Studies*

	Monte Carlo specifications								
MC-study	Heterogenitet in DGP	N	T	B	μ_1	σ_1	θ_c	μ_c	σ_c
I	No	100	250	2500	7	1	4	-	-
II	Yes	100	250	2500	7	1	$\mathcal{N}(\mu_c, \sigma_c)$	4	0.8
III*	Yes	100	250	200	7	1	$\mathcal{N}(\mu_c, \sigma_c)$	4	0.8

*Only study III is estimated assuming heterogeneity.

Note: All studies have $\beta = 0.99$ and $RC = 12$. And 10 chebyshev nodes are used for EV.

estimates $\{\hat{\theta}\}_{b=1}^{B}$ we calculate the Monte Carlo mean:

$$\bar{\hat{\theta}}_B := \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}^{(b)} \qquad B = 2500 \tag{59}$$

The Monte Carlo mean for replacement costs are $\bar{\hat{RC}}_{2500} = 11.93$ indicating a small negative bias since the true $RC = 12$. The same is the case for the estimator of the cost coefficient $\bar{\hat{\theta}}_c = 3.97$ with the true value $\theta_c = 4$.

For $B \to \infty$ the Monte Carlo mean converges in probability to $E[\hat{\theta}]$, hence the bias $E[\hat{\theta}] - \theta_0$ is indicated by $\bar{\hat{\theta}}_B - \theta_0$. In figure 1 we show the cumulative Monte Carlo mean $\bar{\hat{\theta}}_B := \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}^{(b)}$ for $B = 1, ..., 2500$ for $\hat{RC}$ and $\hat{\theta}_c$. The bias can be read off from the figure as the difference between the black non-dotted lines - indicating the true parameter value the cumulative Monte Carlo mean. The negative bias could be due to small sample $N = 100$ or perhaps due to approximation of $EV(.)$ with chebyshev interpolation.

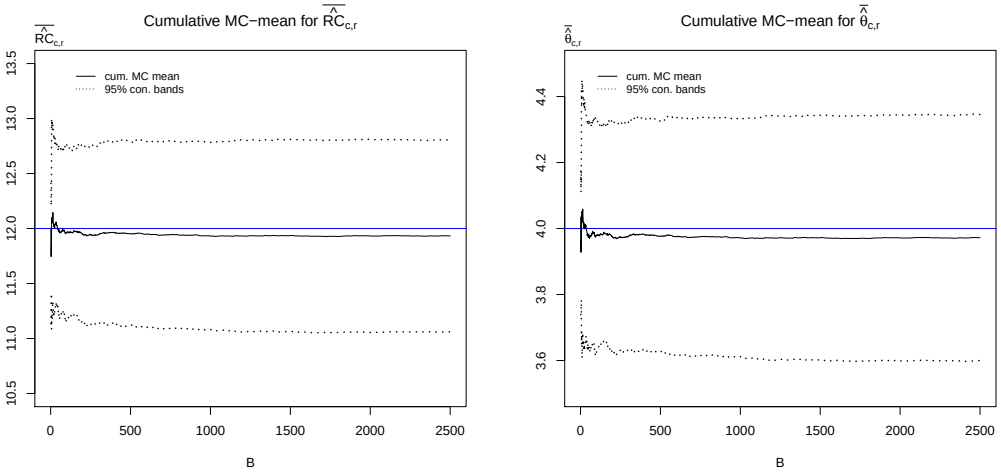


Figure 1: Cumulative MC-means

However while both estimators are slightly negatively biased, the bias is very small if measured in Monte Carlo standard errors found to be 0.44 for $\hat{RC}$ and 0.19 for $\hat{\theta}_c$. This is also apparent from figure 1 where the dotted lines represents a 95% confidence band. In figure 2 we show the empirical standardized quantiles of $\{\hat{\theta}\}_{b=1}^{B}$ for RC and θ_c plotted against the theoretical quantiles of the standard normal distribution. Both estimators are based on the plots approximatedly normal.

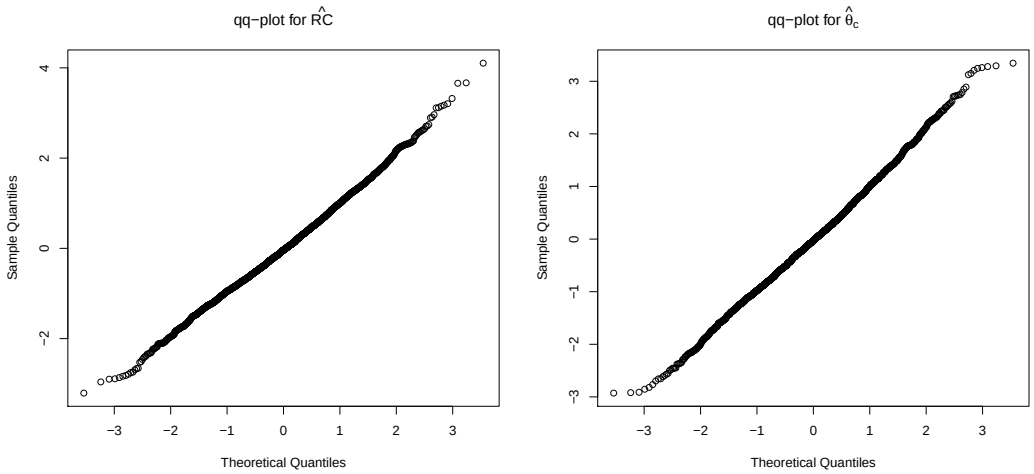


Figure 2: Quantile-Quantile plot for estimators

MC-study II

In the second study we use the full maximum likelihood estimator assuming there is no heterogeneity on data sets generated with heterogeneity. With heterogeneity there is no true value for $\theta_c = \mu_c + \sigma_c z_n$ with $z_n \sim \mathcal{N}(0,1)$ so we compare the estimates of θ_c with the mean value μ_c of the cost coeffiecient. For RC the true value is still 12 but the estimates are very biased with a Monte Carlo mean of 10.45. Likewise the estimator $\hat{\theta}_c$ is a biased estimator of μ_c giving an mean estimate of 3.27 where μ_c was equal to 4. In a static discrete choice model with heterogeneity

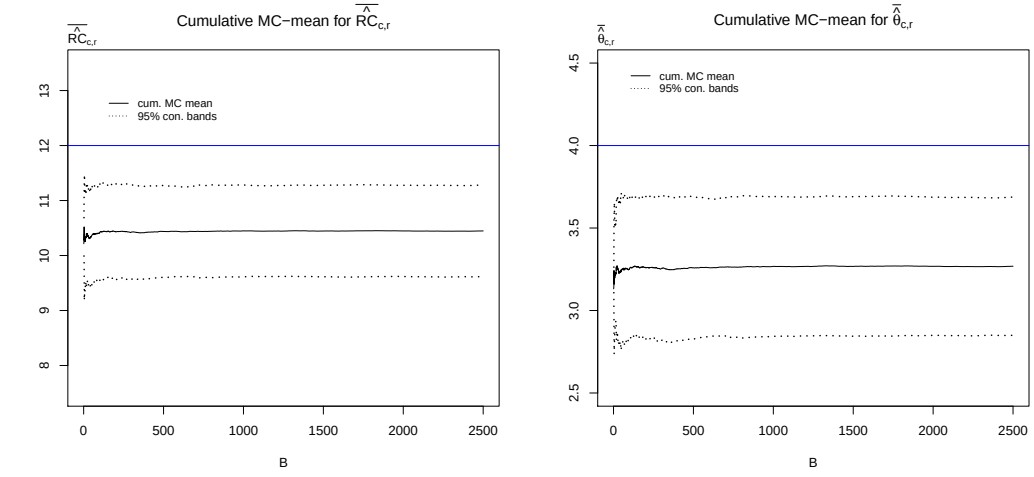


Figure 3: Cumulative MC-means

and random utility linear in the paramaters ignoring the heteroscedasticity will in in general still allow us to estimate the mean of the coefficients up to a scale factor (Scherning 2007: 32). In the present case dividing the estimate of θ_c with the estimate of RC equals 0.31 compared with 0.33 for the same division carried out for MC study I.

MC-study III

In the third Monte Carlo study we generate data with heterogeneity and estimate under the assumption of heterogeneity. The estimation procedure is more time demanding and hence we only produce a 200 estimates. The estimates for replacement costs, mean and the standard deviation of the cost coefficient all appear consistent judged by figure 4. There is no indication of any bias in the estimates given in table 6. The quantile-quantile plots in figure 5 indicates that the estimator is asymptotically normal. The estimator seems overall to be very precise and well behaved in the sense that it behaves as expected on the basis of asymptotic maximum likelihood theory.

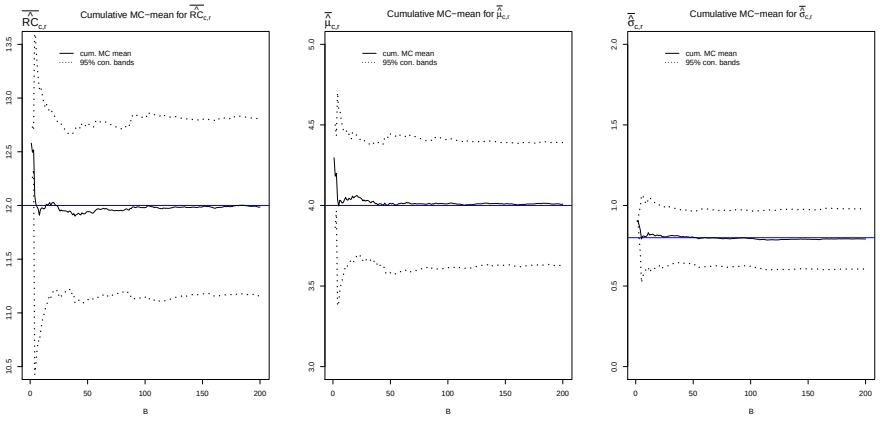


Figure 4: Cumulative MC-means

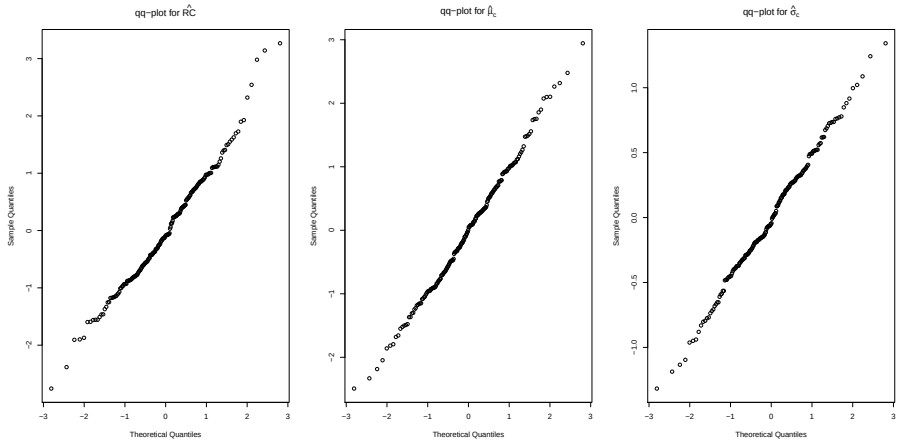


Figure 5: Cumulative MC-means

7 Applying the Function

One of the main benefits of estimating a structurel model is the possibility of doing counterfactual analysis. As in Rust(1987), we can determine the implied demand of an individual using a "bottom-up" approach[7] by aggregating individual decision makers behavior. The object of

[7]referring to the method mentioned in Rust(1987)

Table 6: *Monte Carlo Studies*

	Monte Carlo specifications											
MC-study	$\overline{\widehat{RC}}$	RC	$std\big(\widehat{RC}\big)$	$\bar{\hat{\theta}}_c$	θ_c	$std(\hat{\theta}_c)$	$\bar{\hat{\mu}}_c$	μ_c	$std(\hat{\mu}_c)$	$\hat{\bar{\sigma}}_c$	σ_c	$std(\hat{\bar{\sigma}}_c)$
I	11.93	12	0.445	3.97	4	0.19	-	-	-	-	-	-
II	10.45	12	0.425	3.27	-	0.21	-	4	-	-	0.8	-
III	11.99	12	0.422	-	-	-	4.01	4	0.20	0.79	0.8	0.09

Standard deviation are calculated as the empirical standard deviation of $\{\hat{\theta}\}_{b=1}^{B}$

interest is to compute the annual aggregate demand for replacement of bus engines as a function af the replacement costs RC.

$$\tilde{d}\,(RC) \equiv \sum_{t}^{12}\sum_{i=1}^{M}\tilde{i}_t^m$$

Since the aggregated demand is simply a sum of indicator variables the expected aggregated demand we suggest estimating the expected aggregated demand by simulation. Assuming that the markov proces has a stationary distribution, as is the assumption, simulating the timeseries of considerable length and using only the last observations could plausible be considered random draws from the stationary distribution.

7.1 Simulating the expecting monthly bus engines replacements

The simulation is carried out by creating data according to the procedure in the Monte Carlo studies though with increased amount of time periods to make convergence more plausible. Hence we let $T = 1000$ and $N = 1000$ for the simulated data both with and without heterogeneity.

To calculate the expected monthly average of replacement we simply calculate the fraction of buses with engine replacement in last 12 months of the data, $\frac{\sum_{T-11}^{T}\sum_{n}^{N} d_t^n}{12*1000}$.

In this exercise the interest is whether adding heterogeneity to the cost coefficient has consequences. Hence, will the estimated expected monthly bus engines replacements differ under unobserved heterogeneity. Using the same model setup as in the Monte Carlo I and II, we follow the procedure described above to determine expected annual bus engines replacements:

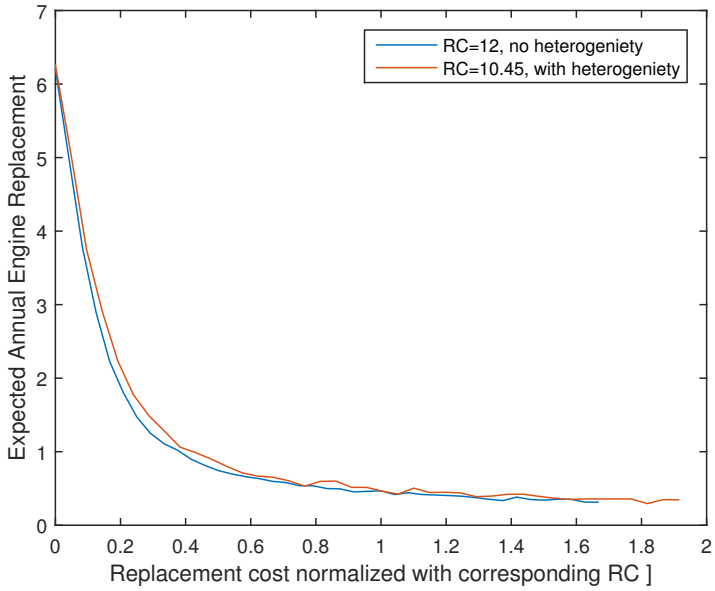
Table 7: Simulate Demand

	Data simulation with model setup as	
RC	Monte Carlo I	Monte Carlo II
	0.039	0.037

Using the simulated approach, we get the result that with unobserved heterogeneity the probability of expected replacement decreases compared to the no heterogeneity. This results has to be taken with huge precaution, as the above estimates above are results of only one simulation procedure for each RC. Having simulated only N buses a deviation on the third decimal is perhaps not to be interpreted as a significant difference. Greater sample size could have been chosen in order to achieve greater precision or alternatively several studies of the same size carried out in order to get an idea of the actual precision.

To derive the implied demand of the individual, the expected monthly bus engine replacements, for different values of RC, for example $\vec{RC} = [0.5; 20]$, we repeat the entire procedure for each coordinate in $\vec{RC}$. The demand curves are shown in figure 6 , where $\vec{RC}$ for each demand curve is normalized with $\hat{RC}$ from is corresponding MC study . Opposed to the Table 7, the

Figure 6: Simulated implied demand



demand curves imply, that the demand curve with heterogeneity has a higher expected monthly bus engine replacement .

7.2 Calculating using Rust(1987) "bottom up" approach

Rust(1987) uses a "bottom up" approach to estimate the expected annual replacement of buses and that the stationary distribution satisfies the equation:

$$\pi(x,i) = \int_y \int_j Pr\left(i|x,\theta\right) p\left(x|y,j\right) \pi_{t,\theta}\left(\mathrm{d}y, \mathrm{d}j\right)$$

Rust(1987) calculates the demand by integrating x out, $\pi_\theta\left(d=1\right) = \int \pi_\theta\left(dx, 1\right)$. Since this is time independent and busses are identical, the annual expected fraction of bus engine replacments is:

$$d\left(RC\right) = 12 \cdot \int \pi_\theta\left(dx, replace\right)$$

Computationally, we need to estimate the unconditional distribution$\pi_\theta\left(x\right)$ and the distribution, $\pi_\theta\left(x|replace\right)$. For this model setup with as following :

1. Create the Markov transition matrix, $Pr\left(d'|x,\theta\right) p\left(x'|x,d\right)$[8]
2. Set an initial transition density of keep, π_0
3. Multiply the the initial distribution with the Markov transition matrix, which returns a new transition density.

[8]For more, see code run_bottom_up_demand.mat

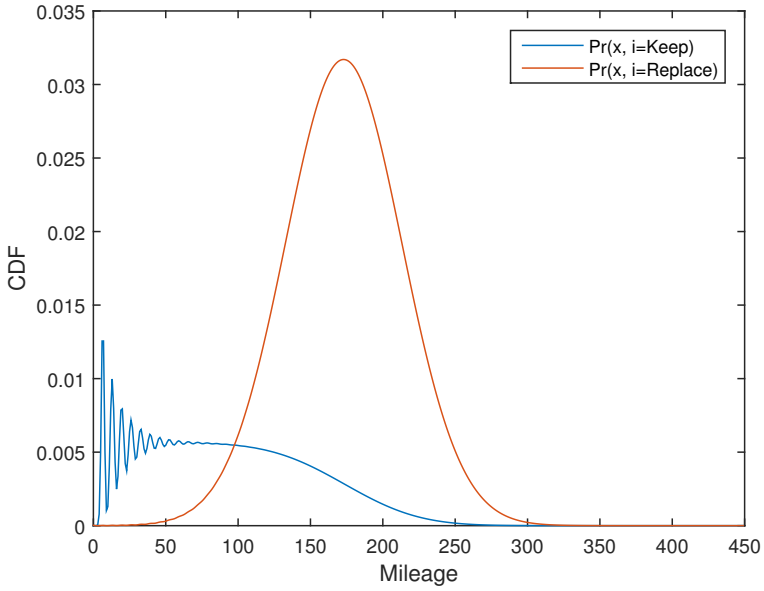
4. Do step 3, now each time with the new transition density until convergence within a tolerance (which returns the stationary transition density)$\rightarrow \pi_\theta(i, x)$

5. $\pi_\theta(x, keep)$ is equal to unconditional density

6. $\pi_\theta(x, replace)$ is equal to unconditional density times $P(replace|x)$

Using the same model setup and parameter values as in the MC study I, the procedure reveals following distributions of $\pi_\theta(x, 1)$ and $\pi_\theta(x, 0)$ in figure:

Figure 7: Distribution of $\pi_\theta(x, 1)$ and $\pi_\theta(x, 0)$



The expected fraction of monthly bus engines replacement is 0.03065. The distribution of the $\pi(x|1)$ appears smooth whereas distribution of the $\pi(x|0)$ suffers of "zig-zags", large for low mileage, but dampens as mileage increases. The behavior in the lower mileage is due a combination of the our transition following the regenerative random walk and discretization. The stochastic term in the transition process is defined as $\mathcal{N}(7, 1)$, hence buses will haven driven 6000 to 8000 in mileage each month with a probability of $68,27$. If a bus is replaced, and thereby resetting mileage to $x = 0$, the bus would have driven 6000 to 8000 in mileage each month with a probability of $68,27$, and a probability of driving yet another 6.000 to 8.000 miles, 12.000 to 16.000 in total, in the next period of $46,61\%$. One approach would be to either lower the average for a more smooth distribution [9]

Obviously, as this method isn't done by simulating, but derived, Harold Zucher's bus data could have been used. The reason we have chosen not to, is that bus types in the fleet with heterogeneity are sorted. Hence, since consequences of unobserved heterogeneity are what of interest, we have chosen to simulate data.

With the distributions determined, we now calculate the expected annual bus engines replacements using the true parameters and ML-estimation from MC-study II:

[9]Mean 7 and std. 1 was chosen to secure non-negative mileage

Table 8: Expected monthly bus engines replacements, "bottom-up" approach

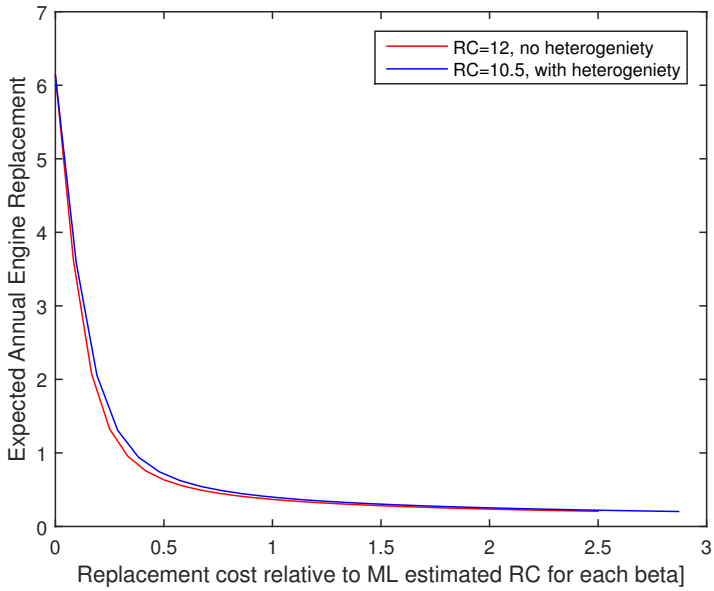
	Bottom up with model setup as	
RC	Monte Carlo I	Monte Carlo II
12	0.03065	0.03099

This result contradicts with the table results in simulated approach but agreeing with the figure in simulated approach, the latter seems to simulate that demand is higher.

These heavily relies on the assumption of the identical busses. As an unobserved heterogeneity imply, the buses won't be identical, as each bus may face different replacement cost.

Again, the purpose of this section is to calculate the implied demand, and again determine if unobserved heterogeneity of our version would have any consequences. As in simulated, we calculate as in the "bottom up" process described above over a grid of RC, $\vec{RC}$, which return the implied demand function, which i illulstrated in figurr ?? again normalizing the $\vec{RC}$ with the correpsonding RC[10]

Figure 8: Demand buttom up approach



The demand curve with ML-estimates is under the implied demand over nearly the entire grid, the demand curve is lower, which seems to be the case with demand curves in the simulations approach.

This section applied the two methods to determine the expected replacement of bus engines and an individuals demand curve. The unobservered heterogeneity may have an impact, but it need further investigation reject or accept that.

[10]The expeted total amount of replacemnt for

$$d\left(RC\right) = 12 \cdot \int_{c}\int_{x} \pi_{\theta}\left(dx, 1, dc\right)$$

8 Conclusion

In this paper we have formulated a version of the bus engine replacement model including heterogeneity in the cost coefficient. We approximated the expected value function using chebyshev nodes and found that not more than 10 chebyshev nodes where necessary in order to achieve satifying precision. The low number of chebyshev nodes reduces computation time compared to techniques using larger number of nodes.

When solving for the expected value function integrals were solved using Gauss Hermite quadrature a technique found to be more sufficient than Monte Carlo draws or equiprobable points. This is hardly surprising giving the choice of a normal distribution for the transition density of the regenerative random walk. However the precision of the techniques with as small a number of nodes as 2, is astonishing.

In the Monte Carlo study of the maximum likelihood estimators we found that ignoring heterogeneity in the cost coefficient lead to a negative bias. The negative bias where present not only in estimates of the cost coefficient, assumed random, but also in the non-random constant for replacement costs.

Using simulation techniques we estimated the expected demand in the Monte Carlo set ups we considered. The results of the approach appear reasonable but should offcourse not be interpreted as a validation of the approach.

9 References

Aguirregabiria, Victor and Pedro Mira (2002) »Swapping the Nested Fixed Point Algorithm: A Class of Estimators for Discrete Markov Decision Models« in *Econometrica* Vol. 70 No. 4 pp. 1519-1543.

Judd, Kenneth L. (1998) *Numerical Methods in Economics.* MIT Press: London, England.

Judd, Kenneth L. & Che Lin Su (2012) »Constrained Optimization Approaches to Estimation of Structural Models« in *Econometrica* Vol. 80 No. 5 pp. 2213-2230.

Rust, John (1987) »Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher« in *Econometrica* Vol. 55 No. 5 pp. 999-1033.

Rust, John (2000) *Nested Fixed Point Algorithm Documentation Manual: Version 6.* http://editorialexpress.com/jrust/nfxp.html

Scherning, Bertel & Fedor Iskhakov, John Rust, Lee Jinhyuk, Kyoungwon Seo (2015) *Constrained Optimization Approaches to Estimation of Structural Model: Comment* Working paper. http://bschjerning.com/research.htm

Scherning, Bertel (2007) *An Estimation Procedure for Mixed Markov Decision Models of Discrete Choices.* Revised draft recieved by author on request.

10 Appendix A: Integration technique results

The tables show the *error* := $100 \cdot \max|\mathbf{p} - \mathbf{p}_0|$ as the max deviance in percentage points of the choice probabilities as explained in section 4.

Table 9: *Gauss Hermite Integration*

	Number of Gauss Hermite nodes						
	2	4	8	16	32	64	100
2	45.58	45.58	45.58	45.58	45.58	45.58	45.58
4	3.22	3.22	3.22	3.22	3.22	3.22	3.22
6	0.31	0.31	0.31	0.31	0.31	0.31	0.31
8	0.83	0.83	0.83	0.83	0.83	0.83	0.83
10	0.05	0.05	0.05	0.05	0.05	0.05	0.05
20	0.03	0.03	0.03	0.03	0.03	0.03	0.03
50	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Note 1: c-nodes are the number of chebyshev nodes

Note 2: Transitionparameters $\mu_1 = 7$ and $\sigma_1 = 1$

Table 10: *Equiprobable Integration*

	Number of nodes					
C-nodes	10	50	100	500	1000	5000
2	45.58	45.58	45.58	45.58	45.58	45.58
4	3.22	3.22	3.22	3.22	3.22	3.22
6	0.31	0.31	0.31	0.31	0.31	0.31
8	0.83	0.83	0.83	0.83	0.83	0.83
10	0.05	0.05	0.05	0.05	0.05	0.05
20	0.03	0.03	0.03	0.03	0.03	0.03
50	0.00	0.00	0.00	0.00	0.00	0.00

Note 1: c-nodes are the number of chebyshev nodes

Note 2: Transitionparameters $\mu_1 = 7$ and $\sigma_1 = 1$

Table 11: *Monte Carlo integration with Halton draws*

	Number of draws					
C-nodes	10	50	100	500	1000	5000
2	45.59	45.59	45.59	45.58	45.58	45.58
4	3.28	3.26	3.29	3.23	3.23	3.23
6	0.33	0.31	0.33	0.30	0.30	0.31
8	0.88	0.86	0.88	0.83	0.83	0.83
10	0.07	0.05	0.08	0.04	0.04	0.04
20	0.05	0.03	0.06	0.02	0.02	0.03
50	0.05	0.03	0.05	0.01	0.01	0.00

Note 1: c-nodes are the number of chebyshev nodes

Note 2: Transitionparameters $\mu_1 = 7$ and $\sigma_1 = 1$

Table 12: *Gauss Hermite Integration*

	Number of Gauss Hermite nodes						
C-nodes	2	4	8	16	32	64	100
2	59.29	59.29	59.29	59.29	59.29	59.29	59.29
4	0.79	0.82	0.83	0.83	0.83	0.83	0.83
6	1.45	1.36	1.37	1.37	1.38	1.38	1.38
8	0.22	0.23	0.23	0.22	0.22	0.22	0.22
10	0.19	0.19	0.19	0.19	0.19	0.19	0.19
20	0.01	0.02	0.02	0.02	0.02	0.02	0.02
50	0.02	0.01	0.00	0.00	0.00	0.00	0.00

Note 1: c-nodes are the number of chebyshev nodes

Note 2: Transitionparameters $\mu_1 = 15$ and $\sigma_1 = 5$

Table 13: *Equiprobable Integration*

	Number of nodes					
C-nodes	10	50	100	500	1000	5000
2	59.29	59.29	59.29	59.29	59.29	59.29
4	0.82	0.83	0.83	0.83	0.83	0.83
6	1.38	1.38	1.38	1.38	1.38	1.38
8	0.23	0.23	0.22	0.22	0.22	0.22
10	0.20	0.19	0.19	0.19	0.19	0.19
20	0.01	0.02	0.02	0.02	0.02	0.02
50	0.01	0.00	0.00	0.00	0.00	0.00

Note 1: c-nodes are the number of chebyshev nodes

Note 2: Transitionparameters $\mu_1 = 15$ and $\sigma_1 = 5$

Table 14: *Monte Carlo integration with Halton draws*

	Number of nodes					
C-Nodes	10	50	100	500	1000	5000
2	59.29	59.29	59.29	59.29	59.29	59.29
4	0.82	0.83	0.83	0.83	0.83	0.83
6	1.38	1.38	1.38	1.38	1.38	1.38
8	0.23	0.23	0.22	0.22	0.22	0.22
10	0.20	0.19	0.19	0.19	0.19	0.19
20	0.01	0.02	0.02	0.02	0.02	0.02
50	0.01	0.00	0.00	0.00	0.00	0.00

Note 1: c-nodes are the number of chebyshev nodes

Note 2: Transitionparameters $\mu_1 = 15$ and $\sigma_1 = 5$

11 Appendix B: Gauss Hermite quadrature

In order to find the expectation of a random variable $f(y)$ where $y \sim \mathcal{N}(\mu, \sigma^2)$ we use the weights $(w_1, ..., w_H)$ and the nodes $(z_1, ..., z_H)$:

1. Transform the nodes $(z_1, ..., z_H)$ into $(v_1, ..., v_H)$ using:

$$v_h = \sigma\sqrt{2}z_h + \mu \tag{60}$$

2. Calculate:

$$f_h = f(v_h) = f(\sigma\sqrt{2}z_h + \mu) \tag{61}$$

3. Approximate expectation:

$$E[f(y)] = \int_y f(y)\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{1}{2}\frac{(y-\mu)^2}{\sigma^2}\right)dy \approx \frac{1}{\sqrt{\pi}}\sum_h^H w_h f_h \tag{62}$$