Multi-Level Workforce Planning in Call Centers

Arik Senderovich

Based on MSc thesis supervised by Prof. Avishai Mandelbaum

Industrial Engineering & Management

Technion

August 2012

Outline

1. Introduction to Workforce Planning
   - Definition and Planning Levels
   - Our goal – Multi-Level Model for Call Centers
   - Call Center characteristics for Workforce Planning
2. Multi-Level Workforce Planning in Call Centers
   - Multi-Level Framework – MDP
   - Applying two models to test case Call Center
     - Model Validation
     - Parameter Estimation
   - Results and insights
3. Future Research

Workforce Planning - Definition

A process of aligning workforce capacity with service/production requirements and organizational goals

- Strategic Goals: **Sales, Customer satisfaction**
- Operational Goals: **Waiting times, Abandonment**

Literature Review:

Robbins (2007)

Workforce Planning Levels

Top-Level Models:
Turnover, Promotions, Recruitment

Strategic Planning

Months, Quarters, Years

Planning Periods

Low-Level Models:
Training, Absenteeism, Protocols

Operational Planning

Events, Hours, Days

Top-Level Planning

- **Planning Horizon**: Quarters, Years,…
- **Planning periods**: Weeks, Months,…
- **Control**: **Recruitment** and/or **promotions**
- **Parameters**:
  - Turnover rates (assumed uncontrolled)
  - Demand/Workload/Number of Jobs on an aggregate level
  - Promotions are sometimes uncontrolled as well (learning)
  - Costs: Hiring, Wages, Bonuses etc.
- **Operational regime is often ignored**

Literature Review:

Bartholomew (1991)

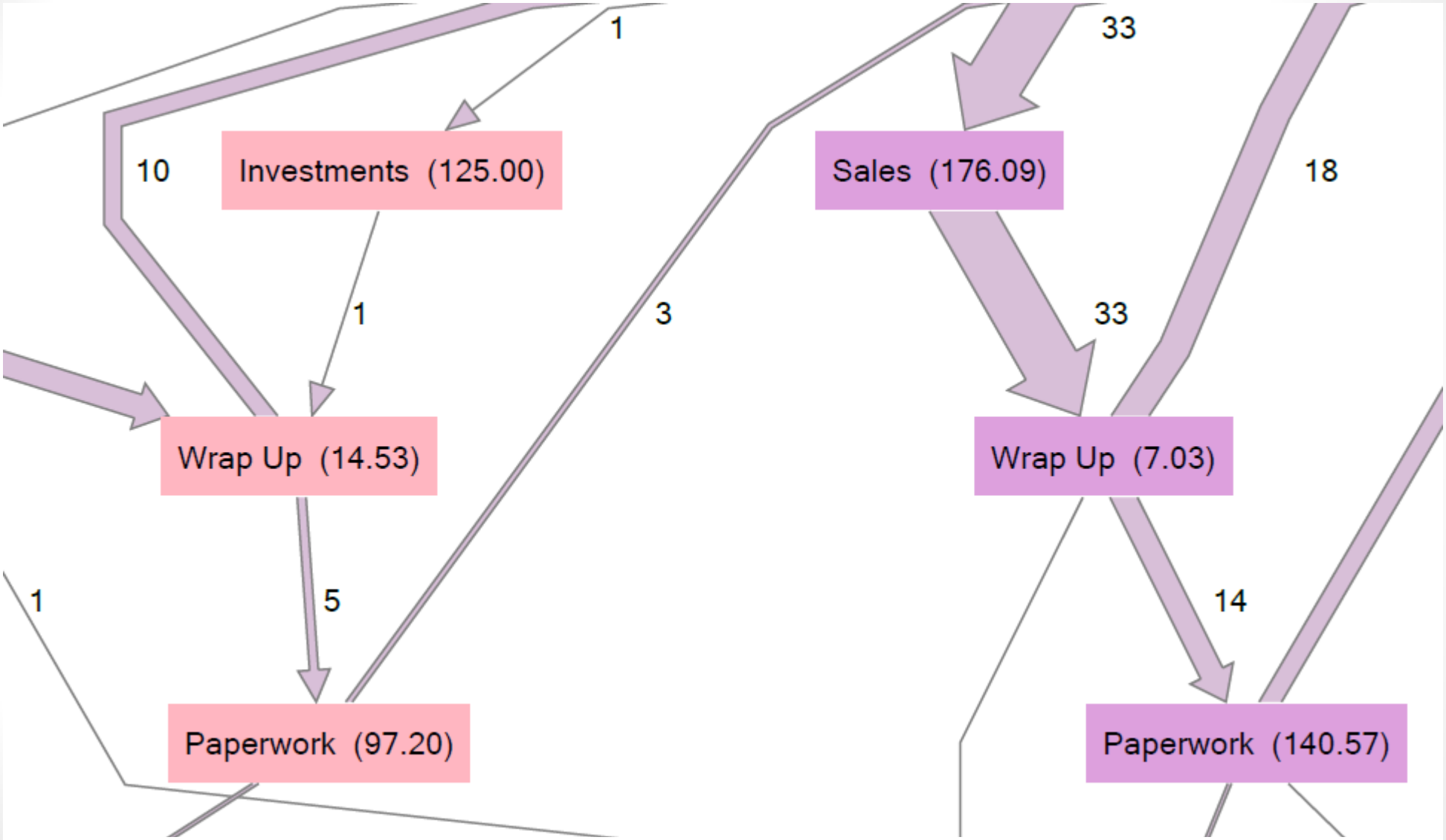
Low-Level Planning

- **Planning horizon**: Months
- **Planning periods**: Events, Hours, Days,….
- **Control**:
  - Daily staffing (shifts, 9:00-17:00,…)
  - Operational regime (work scheduling and routing, managing absenteeism,…)
- **Parameters**:
  - Staffing constraints (shift lengths, work regulations,…)
  - Operational Costs (shifts, extra-hours, outsourcing,…)
  - Absenteeism (On-job, shift)
  - Detailed level demand

Literature Review:

Dantzig (1954); Miller et al. (1974); Pinedo (2010)

Workforce Utilization in Call Centers



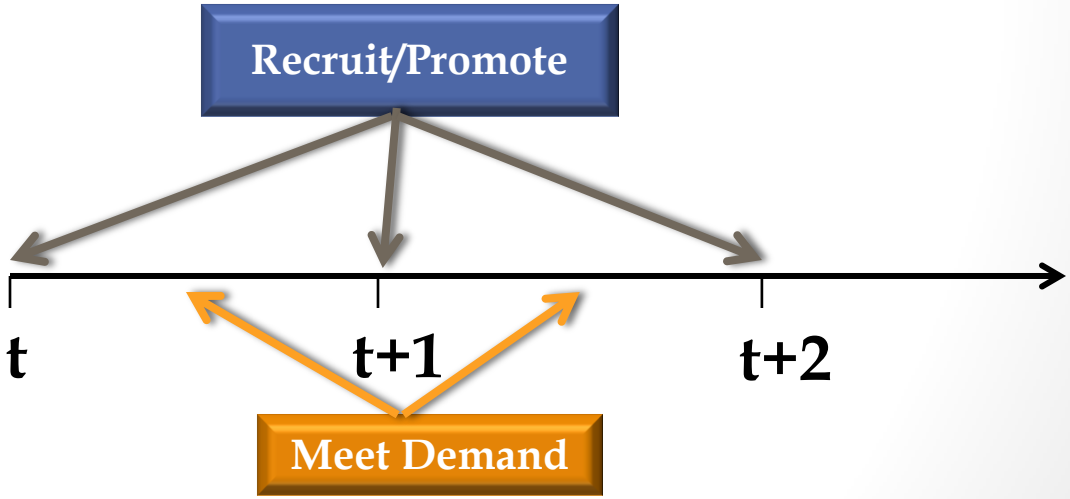
Many customer types, call types, miscellaneous tasks,….

Our goal:

Develop and apply a methodology for Multi-Level Workforce Planning in Call Centers

Multi-Level Planning

- A **single** dynamic model that accounts for **both** planning levels:
  - Low-Level staffing levels do not exceed aggregate constraints
  - Top-Level employed numbers adjusted to meet demand at low-level time resolution
- Dynamic Evolution:



Literature Review:

Abernathy et al., 1973; Bordoloi and Matsuo, 2001; **Gans and Zhou, 2002**

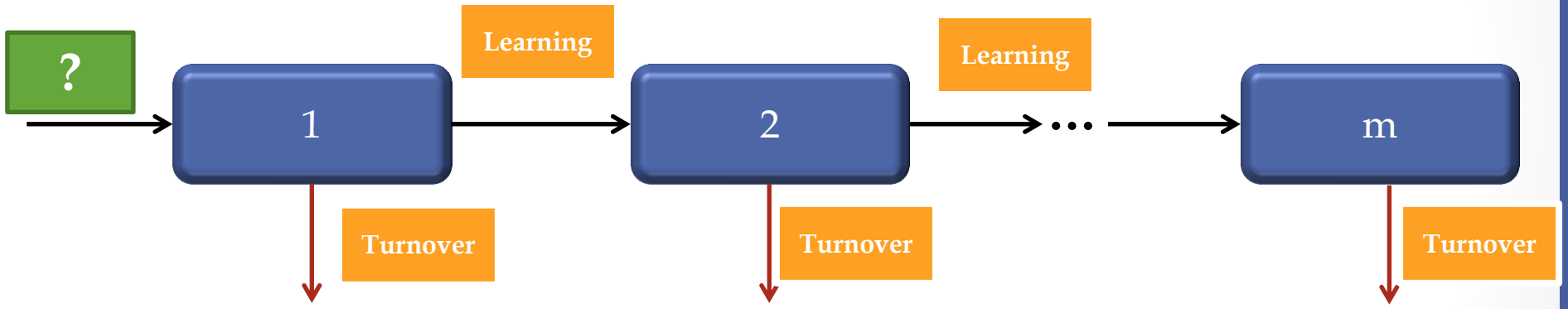
Call Centers – Model Selection

- **High varying demand** (minutes-hours resolution)
- Tradeoff between **efficiency and service level**
- **High operational flexibility** - dynamic shifts
- **Low employment flexibility** - agents learn several weeks
- Multiple skills (Skills-Based Routing)

Proposed models are tested against real Call Center data

Our Framework

- Modeling Workforce Planning in Call Centers via Markov Decision Process (MDP) in the spirit of Gans and Zhou, 2002:



- Control: Recruitment into skill 1
- Uncontrolled: Learning and Turnover
- States *i=1,…,m* may correspond to agent-skills , service speeds or length of service

Model Formulation – Time, State, Control

- T - top-level planning horizon (example: quarters)
- $t = 0,1,...,T$ - top-level time periods (example: months)
- **State space** - workforce at the beginning of period *t:*

$$n_t = (n_{1,t}, n_{2,t}, ..., n_{m,t})$$

- $x_t (\geq 0)$ **- Control** variable at the beginning of period *t*
- **Post-hiring** state-space vector:

$$\tilde{n}_t = (y_t, n_{2,t}, ..., n_{m,t})$$

with $y_t = n_{1,t} + x_t$

- State-space and control are **continuous** (large Call Centers)

Model Formulation – Learning & Turnover

- **Turnover** at the end of period *t*:

$$q_t = (q_{1,t}(y_t), q_{2,t}(n_{2,t}), \ldots, q_{m,t}(n_{m,t}))$$

with

$$q_{i,t}(k) = \tilde{q}_{i,t} k$$

- $\tilde{q}_{i,t}$ - stochastic proportion of agents who turnover

- **Learning** from skill i to i+1, at the end of period *t*, is possible only for those who do not turnover:

$$l_t = (l_{1,t}(y_t), l_{2,t}(n_{2,t}), \ldots, l_{m,t}(n_{m,t}))$$

with

$$l_{i,t}(k) = \tilde{l}_{i,t}(1 - \tilde{q}_{i,t})k$$

- $\tilde{l}_{i,t}$ - stochastic proportion of agents who learn, $\tilde{l}_{m,t} = 0$

Model Formulation - Dynamics

- The system evolves from time *t* to time *t+1*:

$$n_{1,t+1} = (1-\tilde{l}_{1,t})(1-\tilde{q}_{1,t})y_t$$

$$n_{2,t+1} = (1-\tilde{l}_{2,t})(1-\tilde{q}_{2,t})n_{2,t} + \tilde{l}_{1,t}(1-\tilde{q}_{1,t})y_t$$

$$n_{i,t+1} = (1-\tilde{l}_{i,t})(1-\tilde{q}_{i,t})n_{i,t} + \tilde{l}_{i-1,t}(1-\tilde{q}_{i-1,t})n_{i-1,t} \qquad i = 3,\ldots,\mathrm{m}$$

Markov property…

Model Formulation - Demand

- During period t demand is met at low-level sub-periods $s=1,\ldots,S$ (consider half-hours)

- Given J customer types arriving:
  - We define D_t as demand matrix (size $J \times S$)
  - Matrix components are $D_t^{j,s}$:
    - Amount of arriving calls at **time *t*, sub-period *s*** of **call type *j***
    - Example: 10 calls, January 1st , 7:00-7:30, Consulting customer

Model Formulation: Costs

- Low-Level planning is embedded in Top-Level planning in form of an **operational cost function**: $O_t(\tilde{n}_t, D_t)$
- Operational costs considered: **shifting expenses**, **outsourcing** and **overtime**
- $O_t(\tilde{n}_t, D_t)$ is a **least-cost solution** to the Low-Level problem, given period *t* employment levels, recruitment and demand

- Top-Level costs at time *t*:
  - h - Hiring cost of a single agent
  - W_i - Wages and bonuses for skill-level i agents

Model Formulation: Discounted Goal Function

- The discounted total cost that we want to minimize is:

$$\min_{x_0,\ldots,x_T} E\left\{\sum_{t=0}^{T}\left[\alpha^t\left(hx_t + W_1 y_t + \sum_{i=2}^{m} W_i n_{i,t} + O_t(\tilde{n}_t, D_t)\right)\right]\right\}$$

subject to **system dynamics**

- Gans and Zhou: if the operating cost function is jointly convex in $\tilde{n}_t$ there exists an **optimal** "hire-up-to" policy:

$$x_t^* = \begin{cases} y_t^*(n_{2,t},\ldots,n_{m,t}) - n_{1,t} & \text{if } y_t^*(n_{2,t},\ldots,n_{m,t}) \geq n_{1,t} \\ 0 & \text{otherwise} \end{cases}$$

"Hire-up-to" policy - Example

- January workforce – 100 employees
- After turnover – 90 employees
- February demand – 110 employees
- **Myopic** "hire-up-to" – 20 recruits
- Sometimes **NOT** enough considering long-run parameters (demand, flow,…)
- For example: DP dictates hire 30
- If the number is less than 0 then we hire 0

Modeling the Operating Cost Function

- We propose the following model for $O_t(\tilde{n}_t, D_t)$:
  - $w = 1,\ldots,W$ - feasible shifts during time period t
  - $x_{i,w}$ - number of level-i agents staffed to shift w
  - $c_{i,w}$ - cost for staffing level-i agent to shift w

$$O_t(\tilde{n}_t, D_t) = \min_{x_{i,w} \geq 0} \sum_{i=1}^{m} \sum_{w=1}^{W} c_{i,w} x_{i,w}$$

$$\text{s.t.} \quad \sum_{\text{w:I(w,s)=1}} x_{i,w} \geq N_i(D_t^s), \qquad \forall i, s$$

$$\sum_{w=1}^{W} x_{1,w} \leq y_t$$

$$\sum_{w=1}^{W} x_{i,w} \leq n_{i,t}$$

Applying 2 Models to Test Case Call Center

- Models are special cases of Gans and Zhou, 2002
- **Validating assumptions** and **estimating** parameters using real Call Center data
- Comparing results – Models vs. Reality

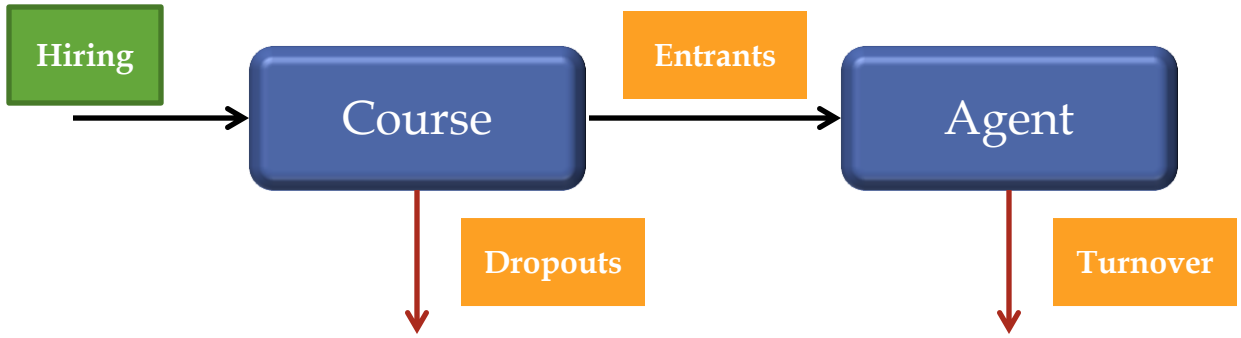
Test Case Call Center: An Israeli Bank

- Inbound Call Center (80% Inbound calls)
- Operates six days a week
  - Weekdays - 7:00-24:00, 5900 calls/day
  - Fridays – 7:00-14:00, 1800 calls/day
- Top-Level planning – quarters
- Low-Level planning – weeks
- Three skill-levels:
  - Level 1: General Banking
  - Level 2: Investments
  - Level 3: Consulting

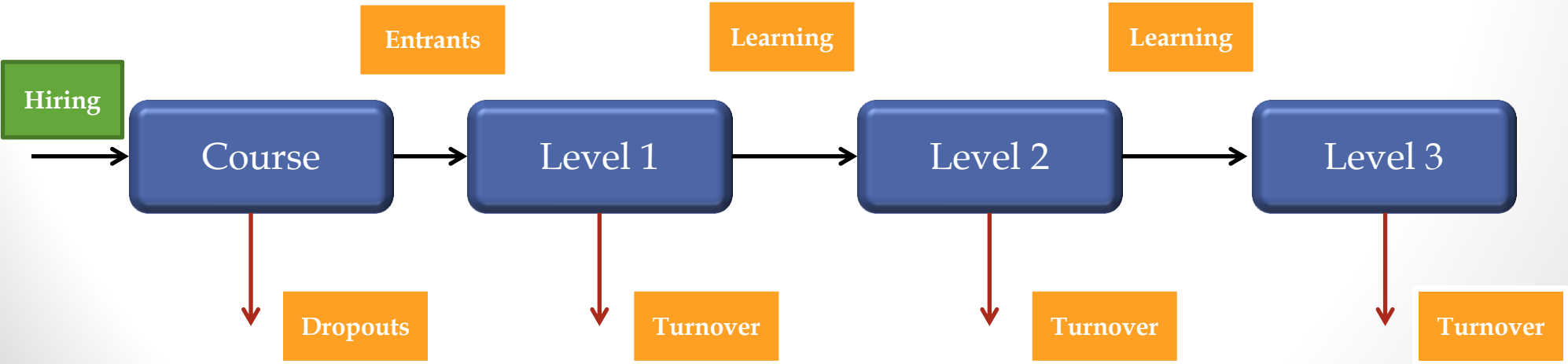
Model Validation and Application

- **Training set:** Year 2010 SEEData + Agent Career data
- **Test set:** Jan-Mar 2011 SEEData
- **Top-Level planning horizon:** 1st Quarter of 2011
- **Top-Level time periods:** Months (January-March 2011)
- **Sub-periods (low-level periods)**: Half-hours

Model1: Base Case Model



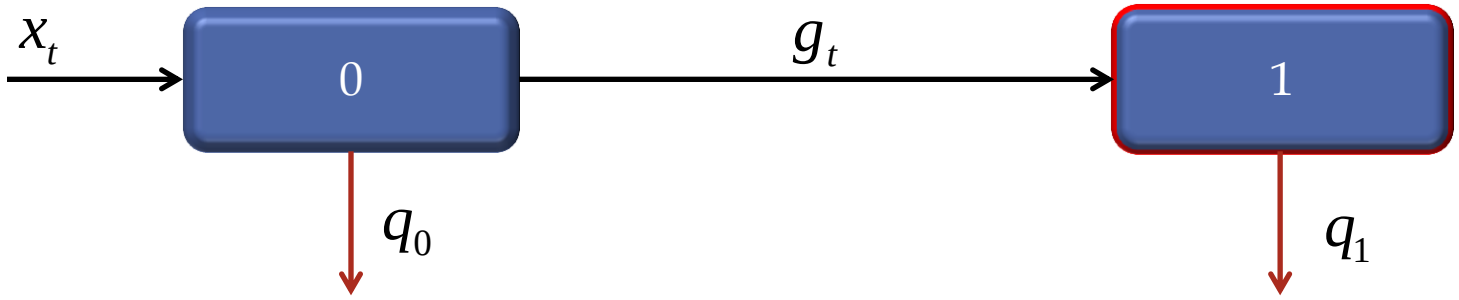
Model2: Full Model



Model 1: Assumptions

- **Single** agent skill (no learning/promotion)
- Deterministic and stationary turnover rate
- Stationary demand
- Recruitment lead-time of one period - **Reality**

Model 1 : Formulation



$$\min_{g_0,\ldots,g_T} \sum_{t=0}^{T} \left[h\frac{g_t}{(1-q_0)} + \overline{W}(g_t + n_t) + O_t(\tilde{n}_t, D_t) \right]$$

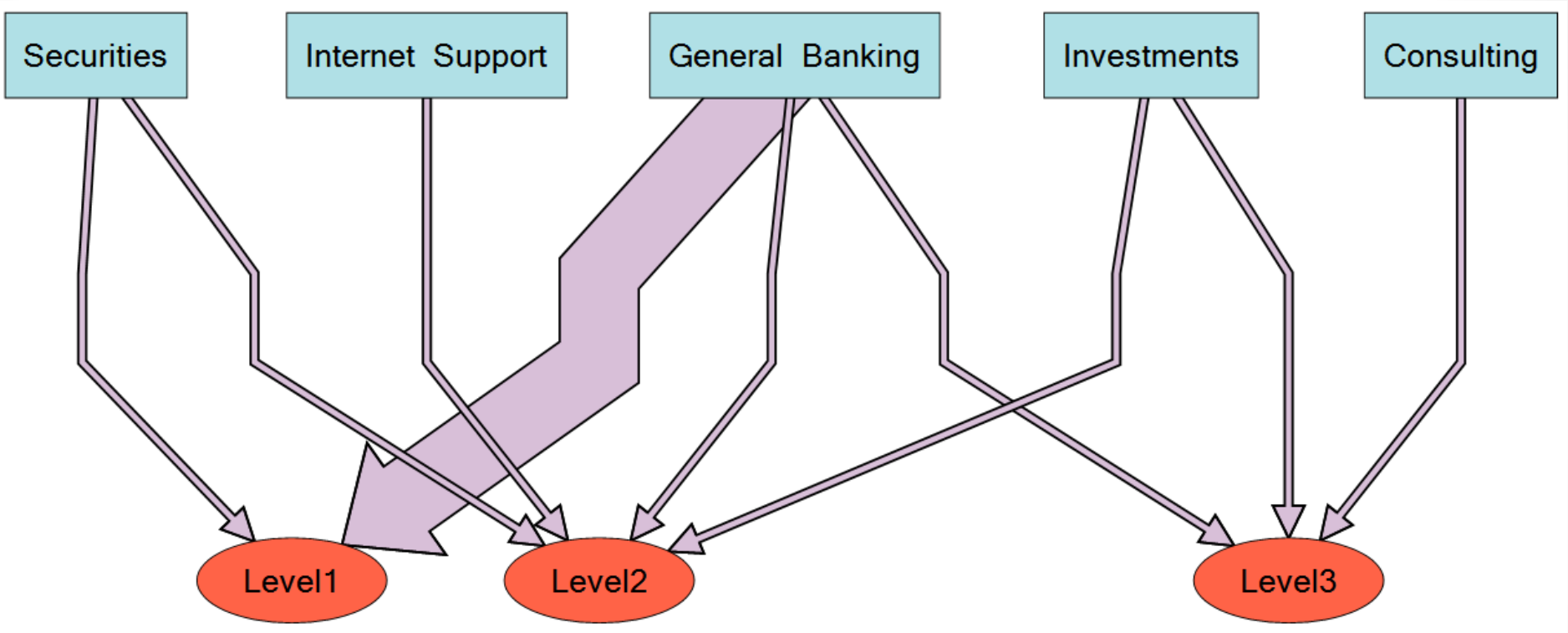
Subject to dynamics:

$$y_t = n_t + g_t$$

$$g_t = (1-q_0)x_{t-1} \geq 0$$

$$n_{t+1} = y_t(1-q_1)$$

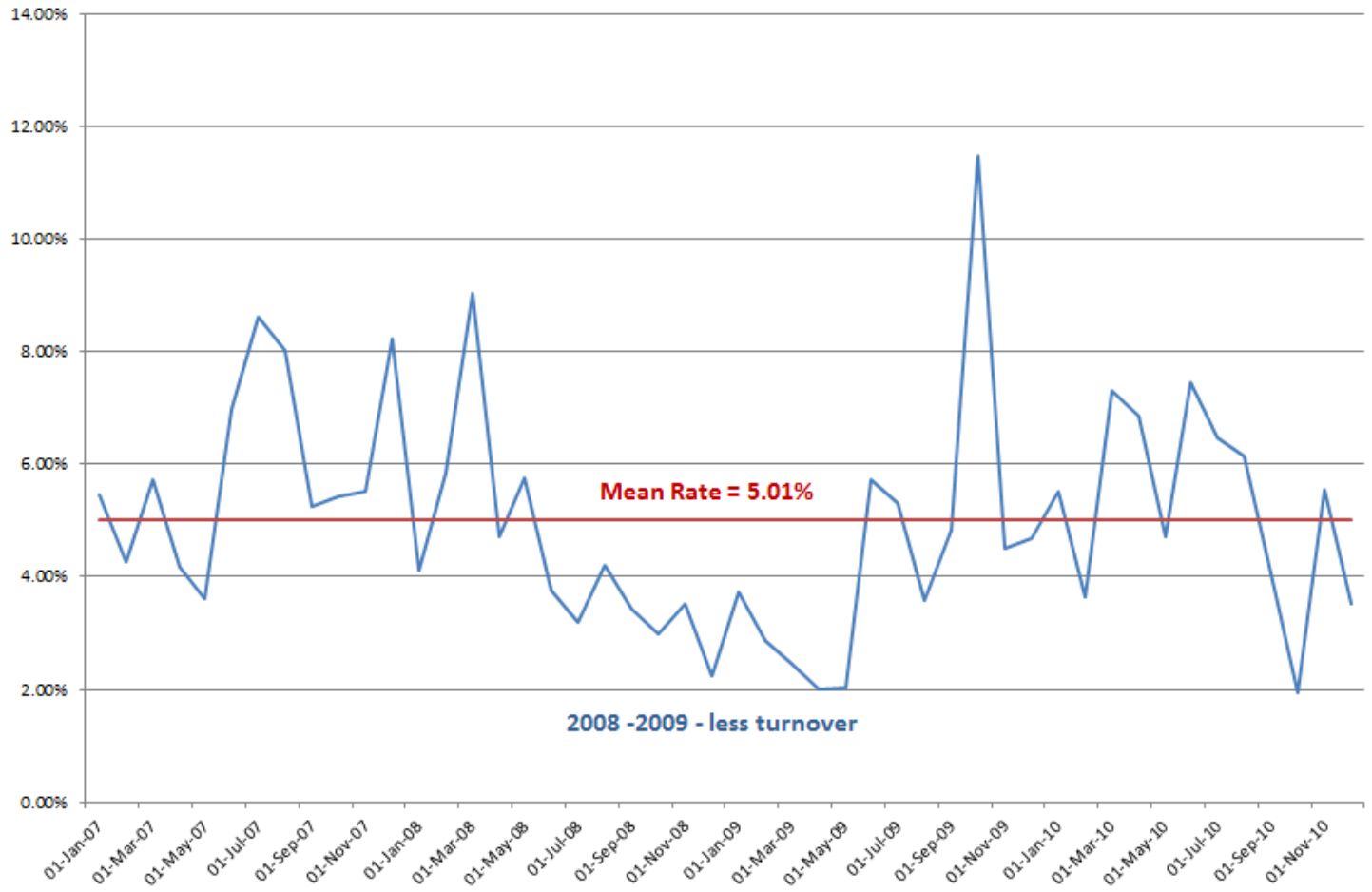
Validating Assumptions: No Learning



No Learning assumption is not valid but Model 1 can still be useful due to simplicity

Validating Assumptions: Turnover

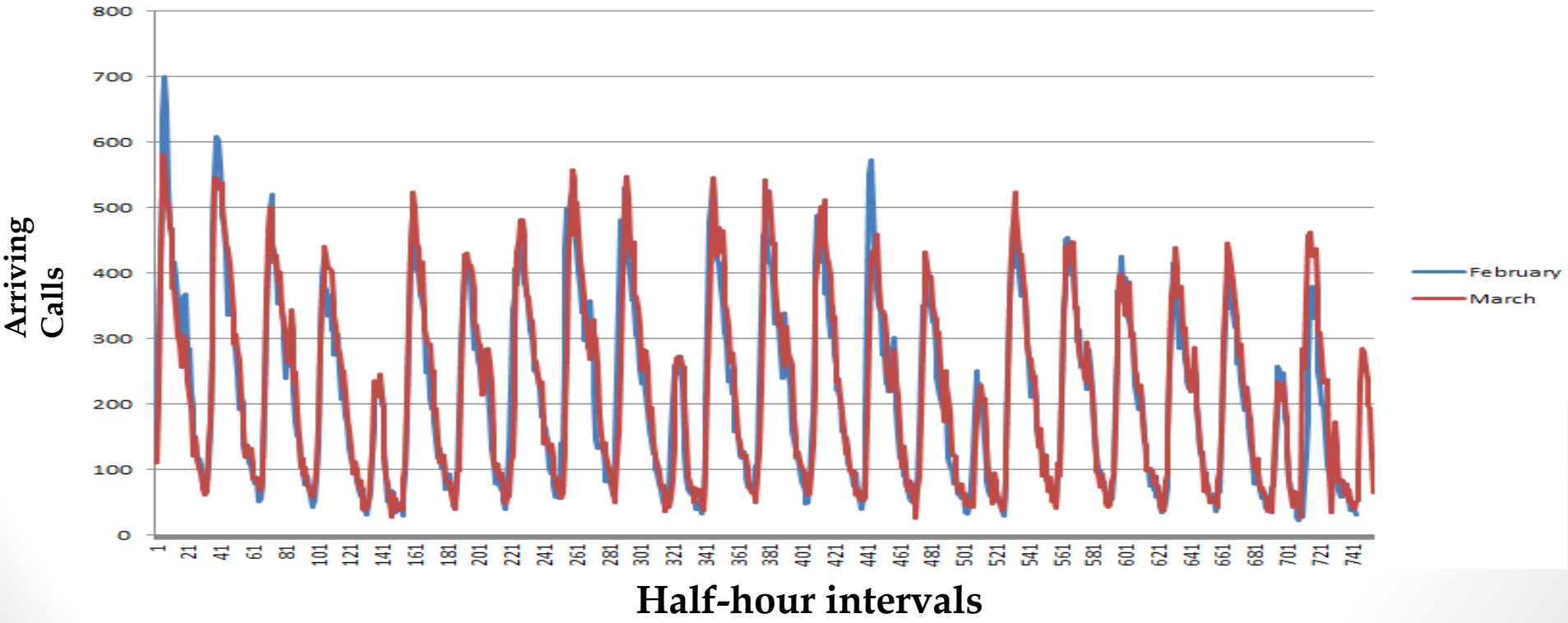
Monthly turnover rate (2007-2010):



Average turnover rate of 2010 serves estimate – 5.27%

Validating Assumptions: Stationary Demand

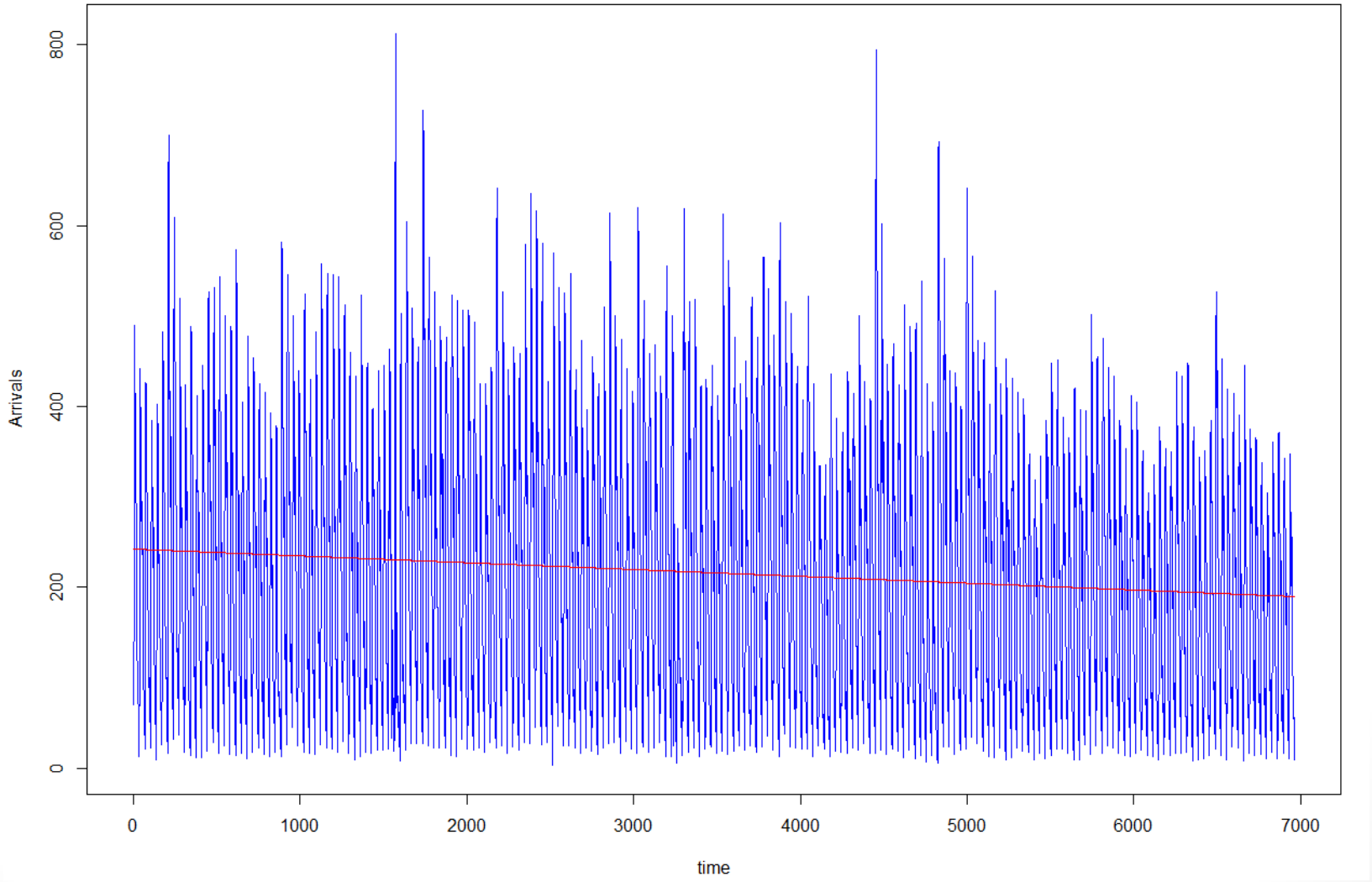
- Demand in half-hour resolution:
  - Not too long - Capturing variability
  - Not too short – Can be assumed independent of each other
- Comparing two consecutive months in 2010, for **total** half-hour arriving volume:



Stationary demand is a reasonable assumption

Validating Assumptions: Stationary Demand

- We now examine the half-hours for entire year 2010:



Model 1: Low-Level Planning

$$O_t(y_t, D_t) = \min_{x_w \geq 0} \sum_{w=1}^{W} \bar{c}_w x_w$$

$$\text{s.t.} \quad \sum_{w:I(w,s)=1} x_w \geq N(D_t^s), \qquad \forall s$$

$$\sum_{w=1}^{W} x_{1,w} \leq y_t$$

Modeling Demand

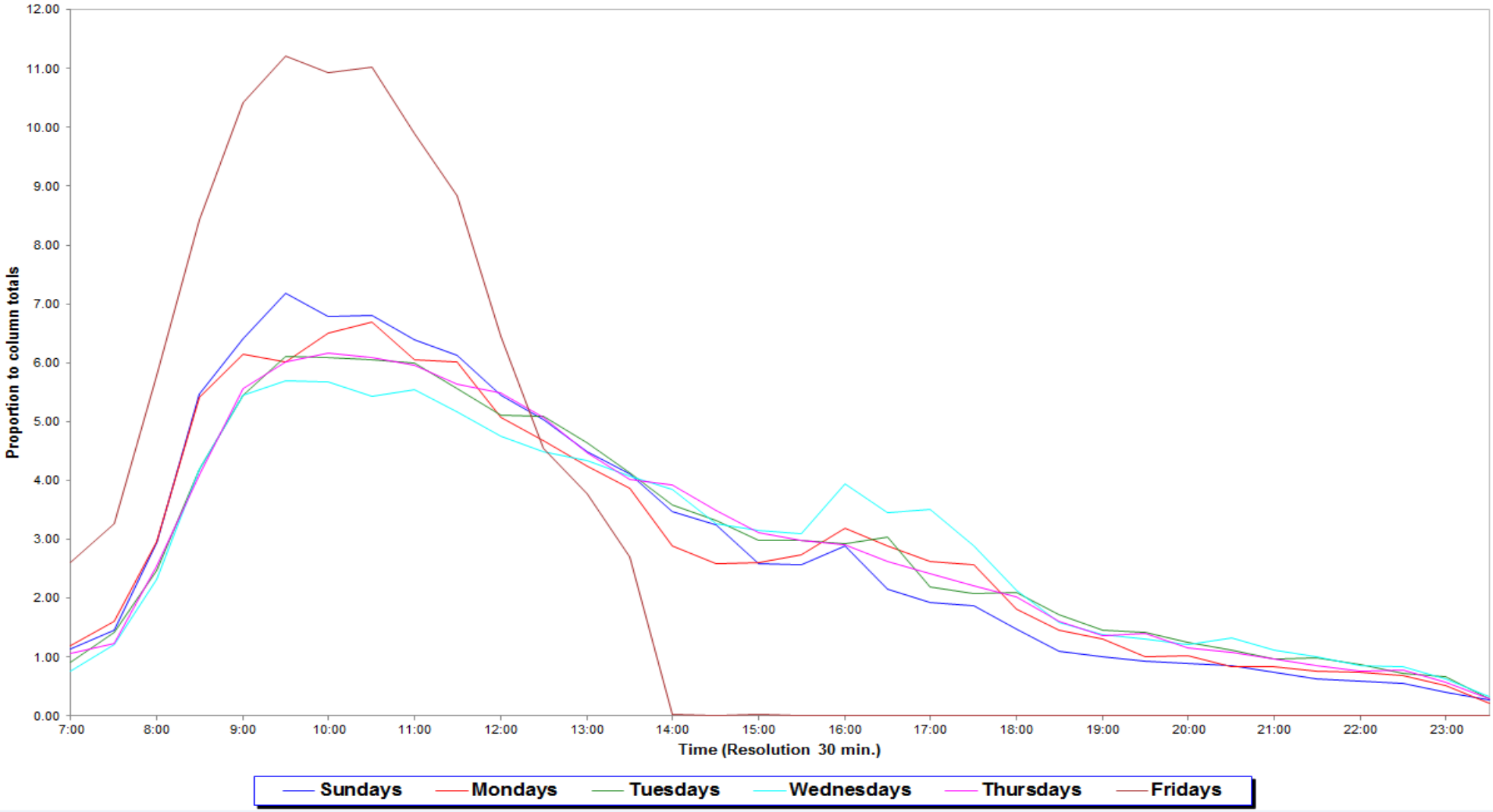
- **General additive model (GAM)** was fitted to demand of October-December 2010 (Hastie et al., 2001):
  - Demand influenced by two effects: **Interval** effect and **Calendar day** effect

$$D_t^{s,c} = \alpha_s + \gamma_c + \varepsilon_{s,c}$$

  - Fitting GAM for each customer class j did not influence results

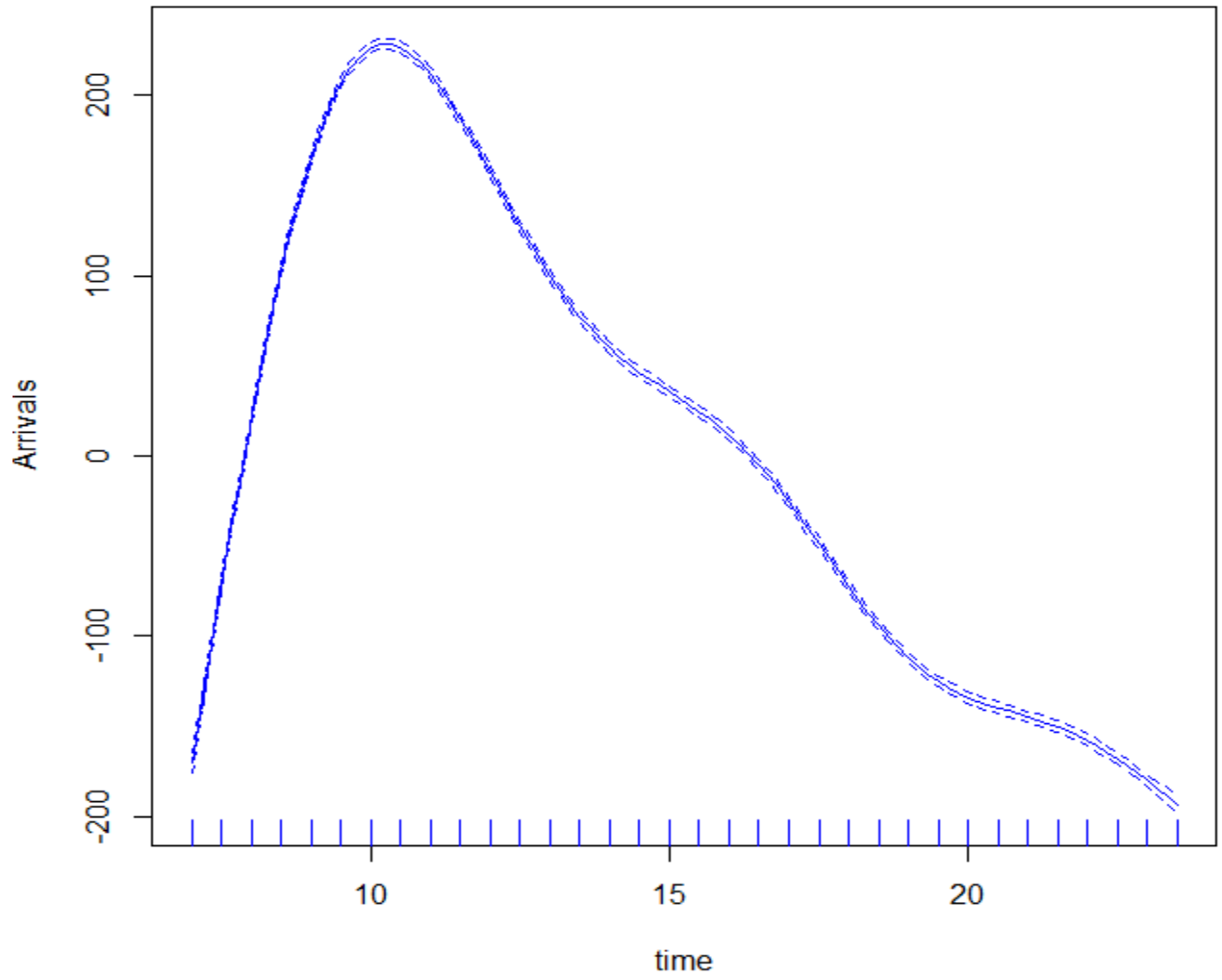
Forecasting demand in Call Centers - Aldor-Noiman et al., 2008

Modeling Demand – Weekdays and Fridays

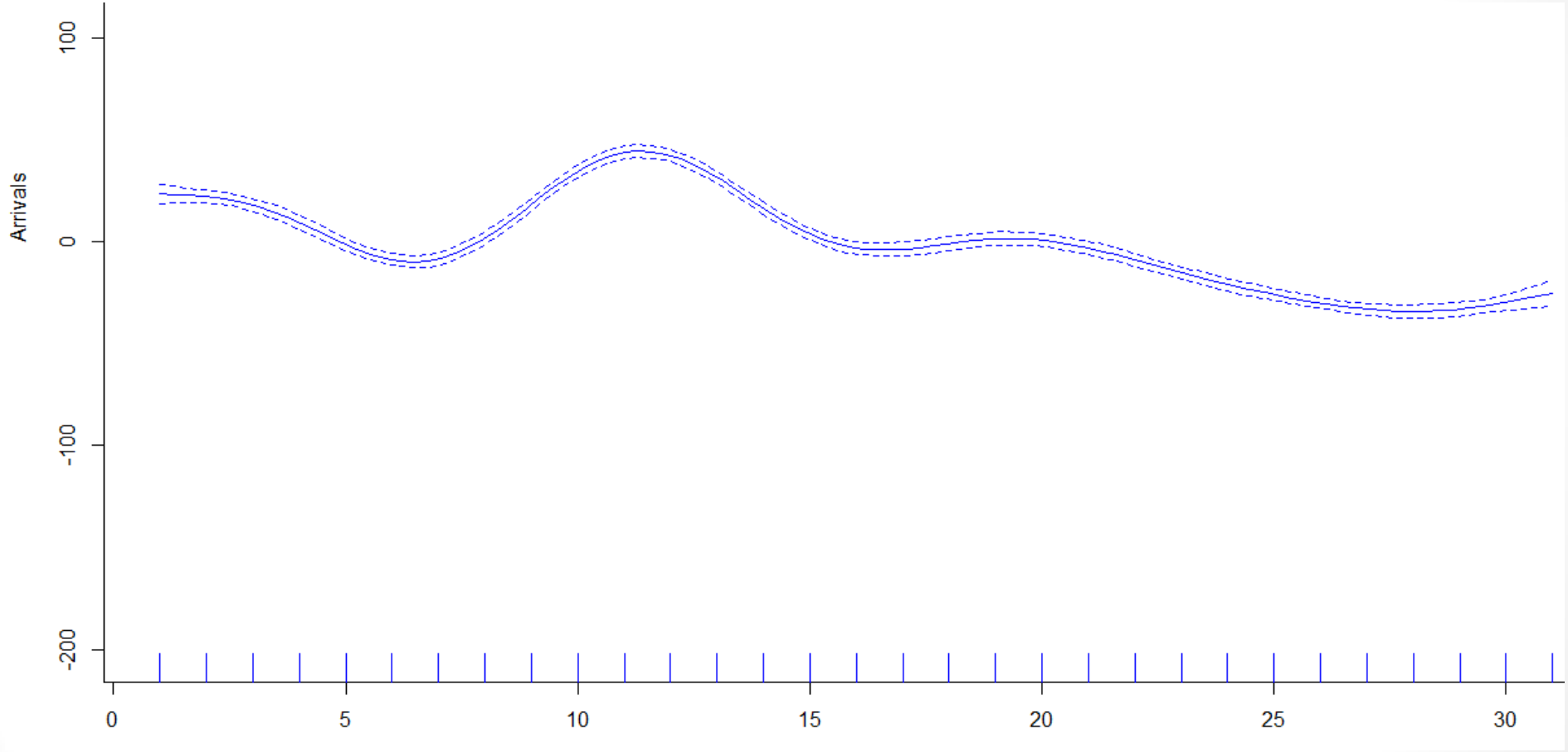


Weekdays effect was not significant for total demand

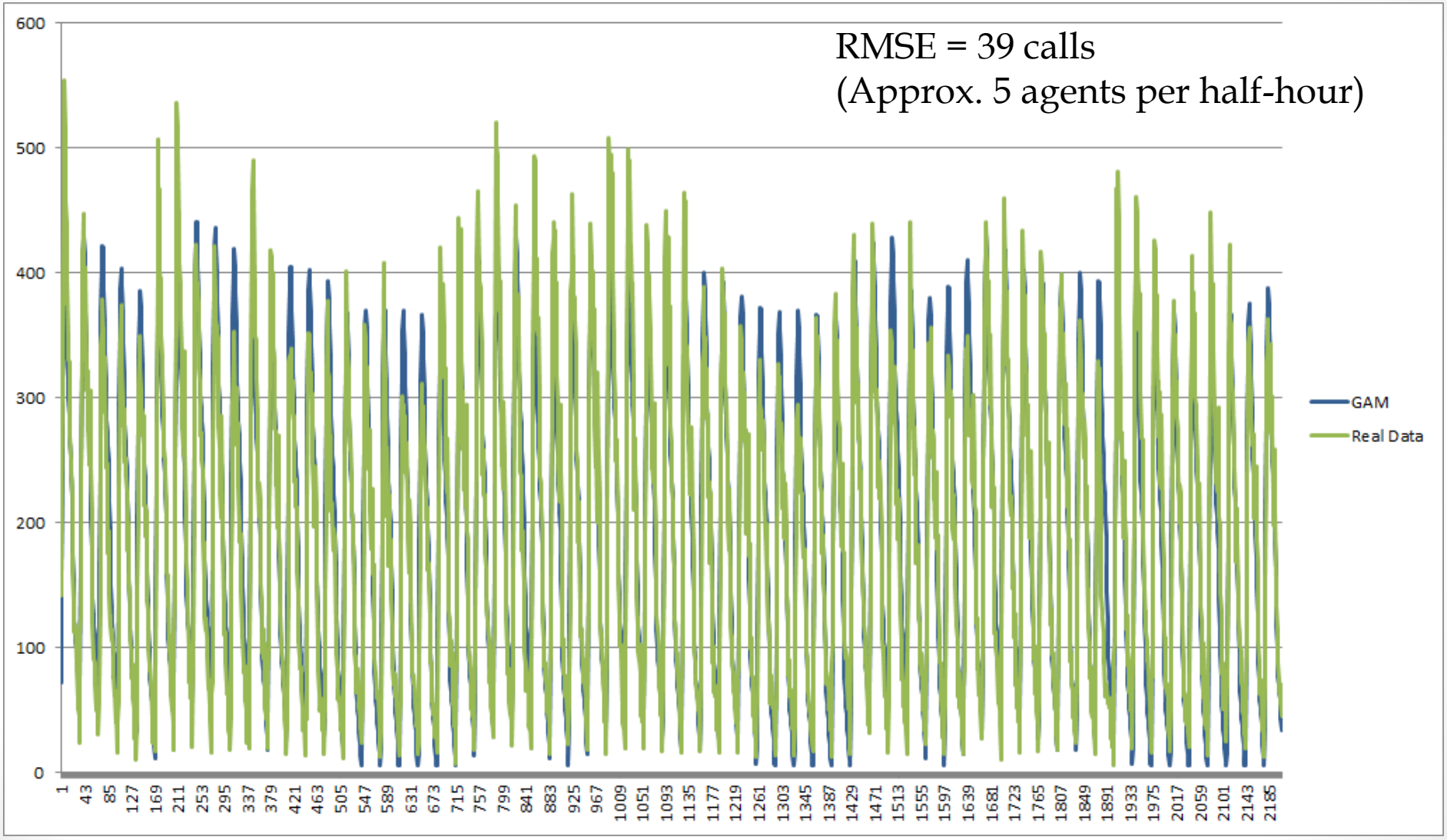
Modeling Demand - Weekday Half-Hour Effect



Modeling Demand - Calendar Day Effect

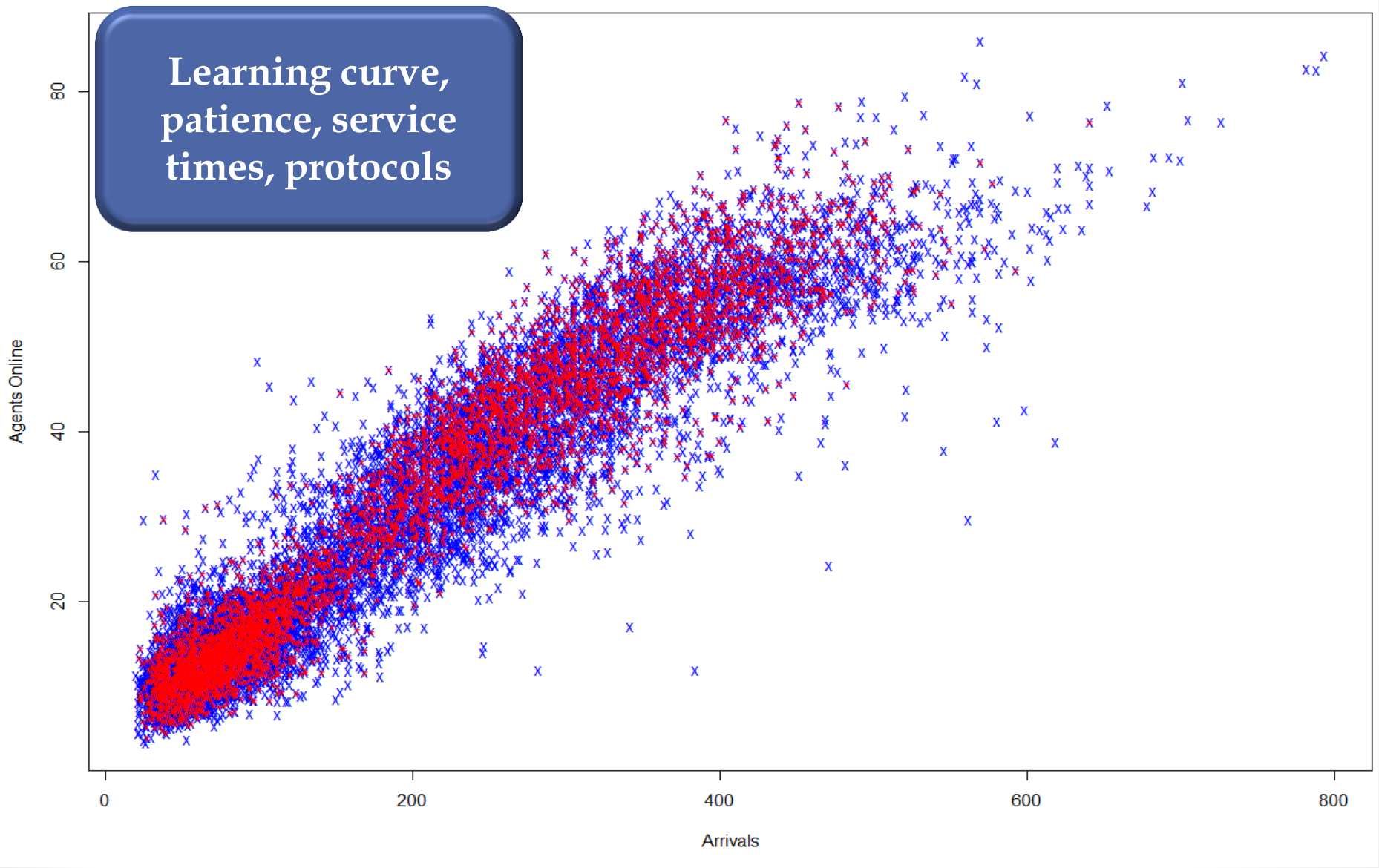


Modeling Demand – Goodness of Fit



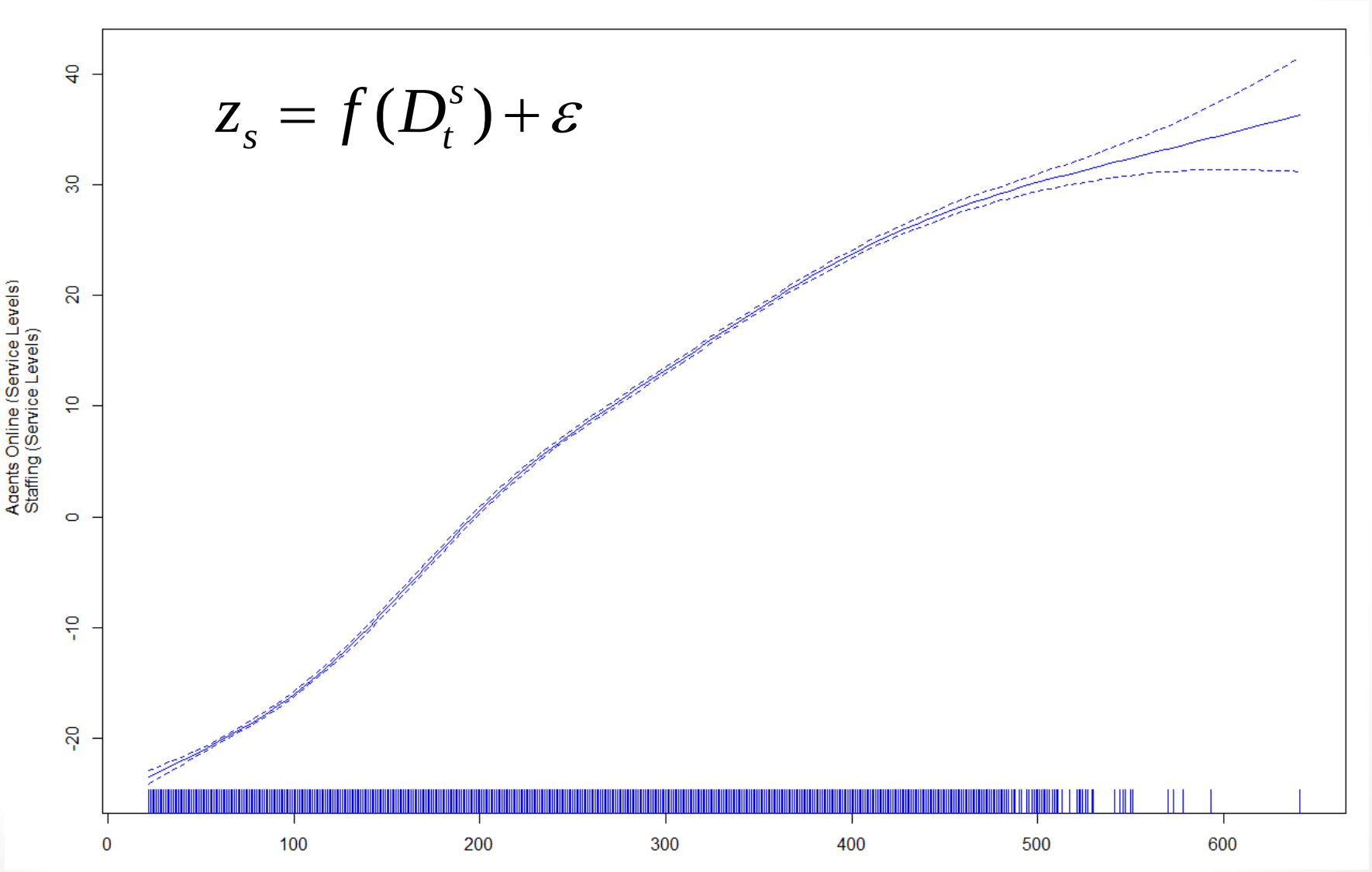
Not much better than fitting whole (de-trended) year 2010

Agents Online – Learning From Data



Staffing Function – Non-linear Spline

$$z_s = f(D_t^s) + \varepsilon$$

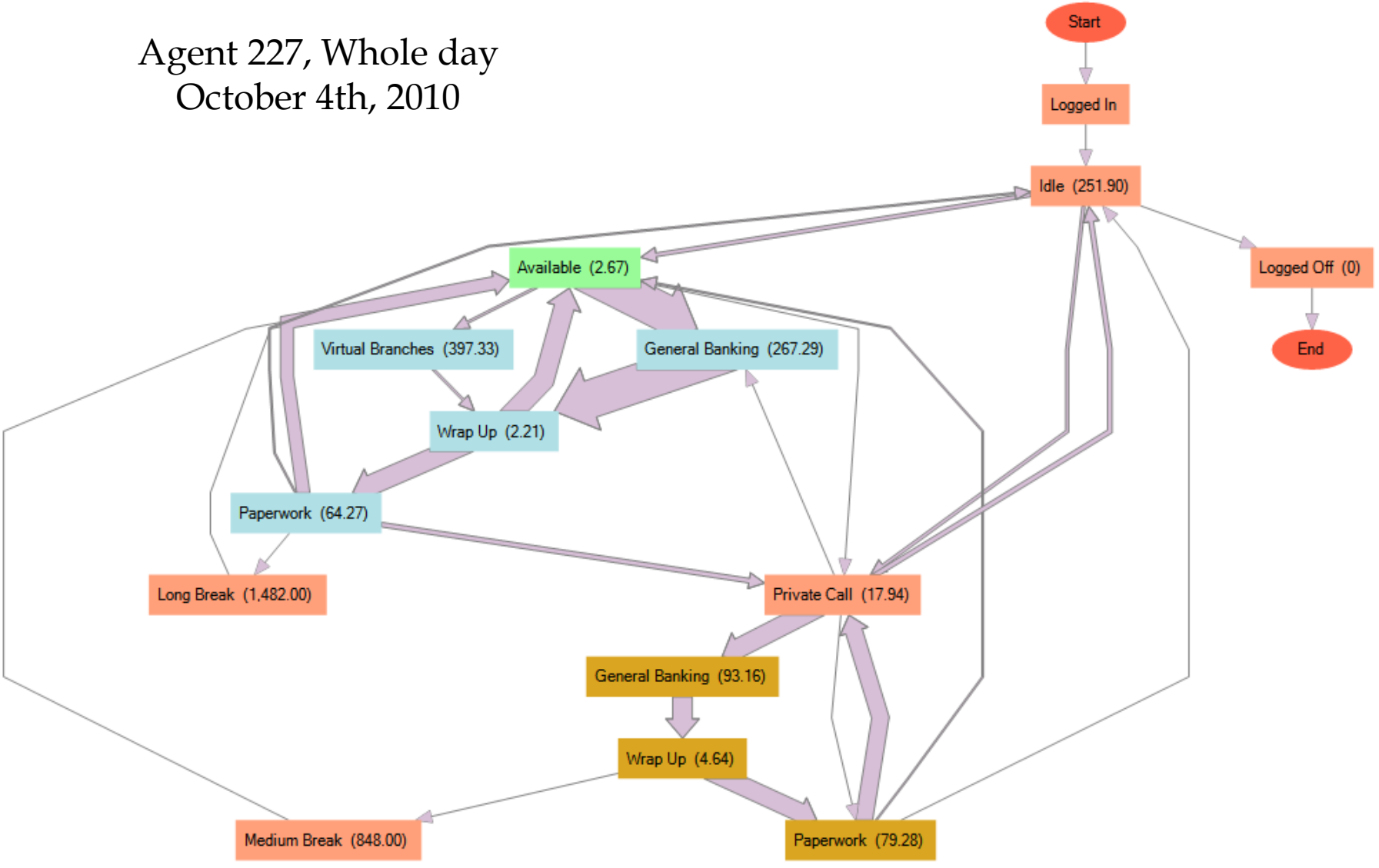


On-job Absenteeism

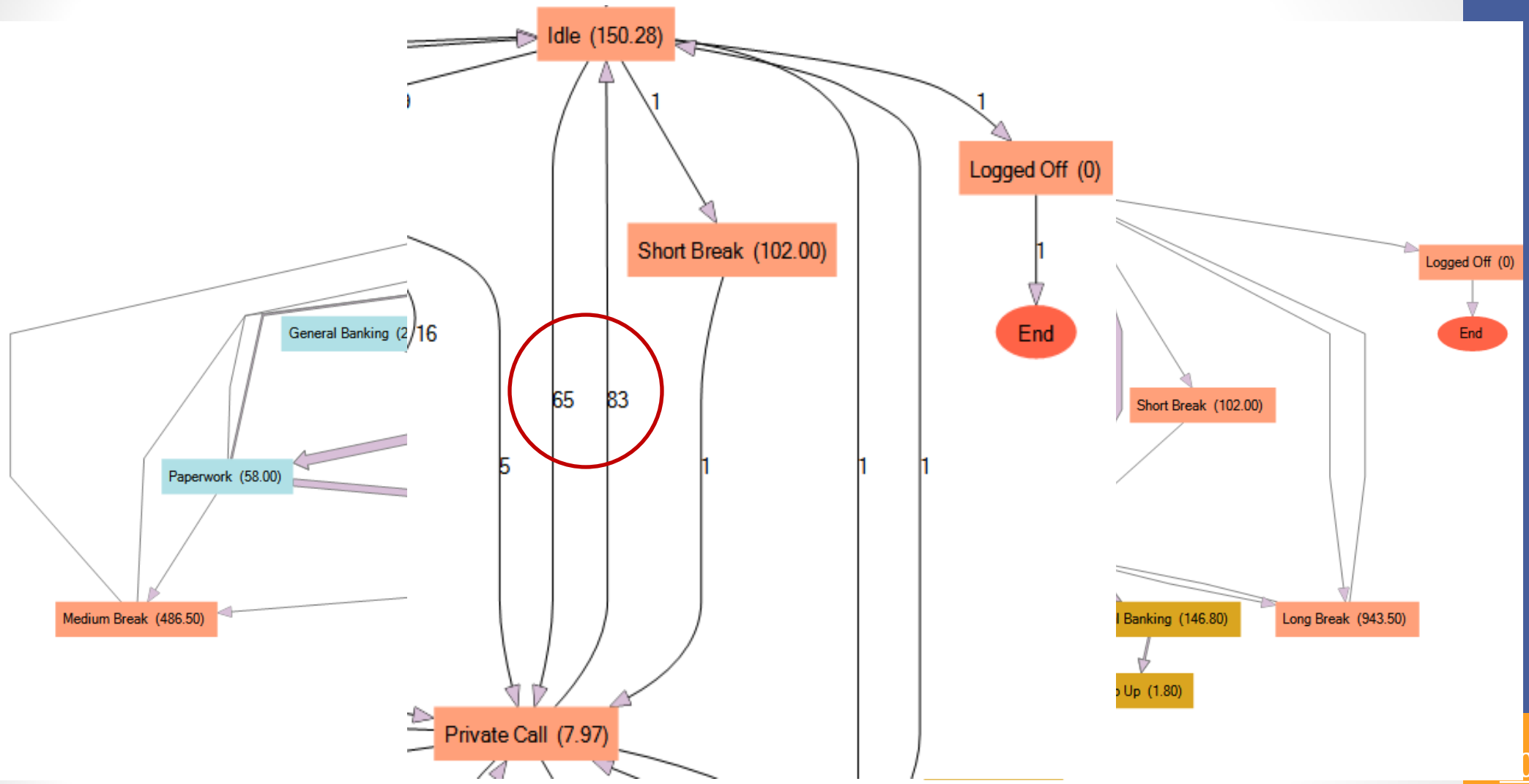
- **During shifts**: agents go on breaks, make outgoing calls (sales, callbacks) and perform miscellaneous tasks
- More (half-hour) staffing is required
- Israeli bank policy:
  - Only **breaks** and some **miscellaneous tasks** are recognized
  - Outgoing calls and other back-office work are important, but assumed to be postponed to "slow" hours
  - Factor of 11% compensation at Top-Level workforce (uniform over all shift-types, daytimes etc.)
- We model absenteeism at low-level resolution and show that it is time varying (great influence on planning)
- We use **Server Networks** to answer questions on agent utilization profile

Newly hired agent

Agent 227, Whole day
October 4th, 2010



Old timer



Agent 513, Whole day
October 4th, 2010

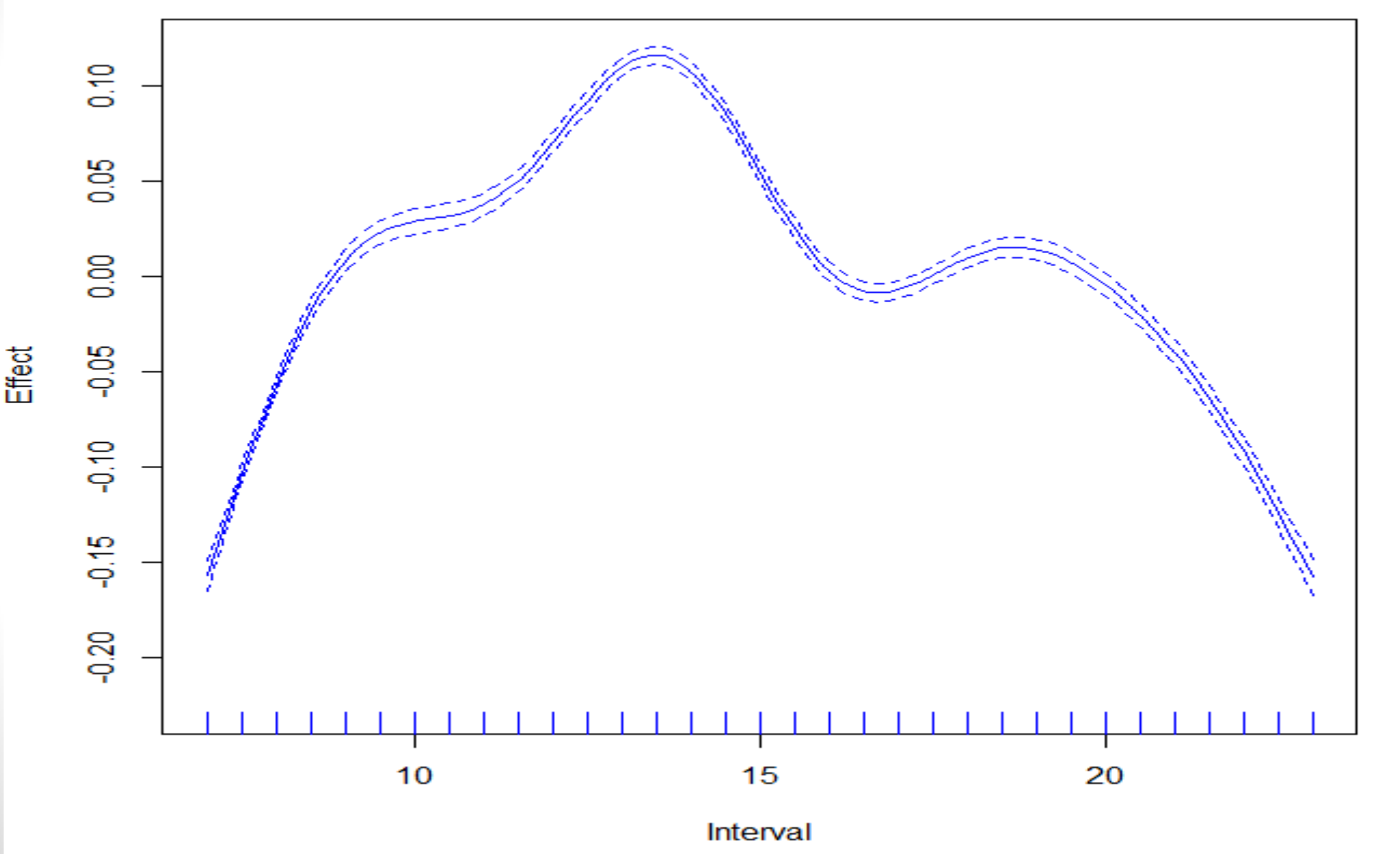
Defining and Modeling Absenteeism

- Absenteeism **rate** per interval *s* as:

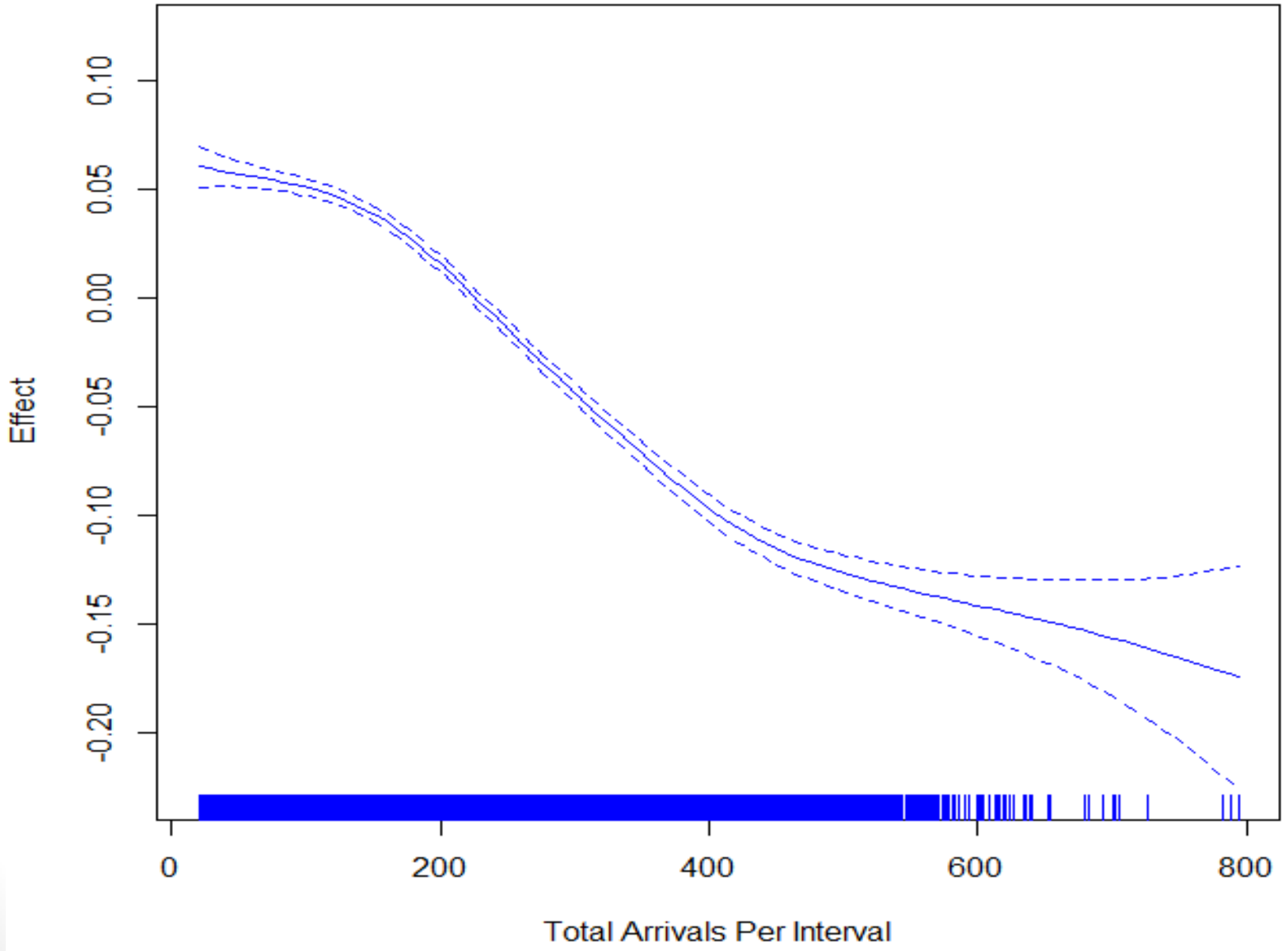
$$p_s = \frac{\text{Total absenteeism per interval}}{\text{Total staffing per interval}} = \frac{a_s}{z_s + a_s}$$

- Absenteeism is defined as breaks and other productive work (management decision)
- GAM model is fitted (again) to **absenteeism rate** with covariates:
  - Time of day
  - Total arrivals per period
- Weekdays-Fridays are separated again
- **On-shift absenteeism**: between 5% and 35% (average of 23% vs. 11% bank assumption)

Fitting Absenteeism – Time of Day



Fitting Absenteeism – Arrivals



Shift Absenteeism

- Shift absenteeism: agent scheduled to a certain shift and does not appear (health, AWOL,…)
- We model it as probability of not showing up for shift given scheduling
  - No supporting data, thus assuming 12% overhead corresponding to bank policy
  - Given data parameters can be estimated and plugged into operational cost function

Low-Level Planning: Staffing

$$N(D_t^s) = z_s + a_s = \frac{z_s}{(1 - p_s)}$$

$$O_t(y_t, D_t) = \min_{x_w \geq 0} \sum_{w=1}^{W} \bar{c}_w x_w$$

$$\text{s.t.} \quad \sum_{w:I(w,s)=1} x_w \geq \hat{N}(D_t^s), \qquad \forall s$$

$$\sum_{w=1}^{W} x_{1,w} \leq y_t$$

Model 1: Multi-Level Solution

- **Myopic** single-stage "hire-up-to" policy is optimal:
  - Low-Level planning **sets** number of employees for each time period *t*
  - **Gaps** are known in advance and filled
  - Recruitments are made one period **ahead**
- **Example:**
  - Low-Level solution January 2011 is 100 employees
  - In the beginning of December we have 100 employees
  - We know that 10 will turnover at the end of December
  - We hire in December 10 to replace them (if no dropouts occur)

Model 2 : Assumptions

- Model 1 is extended to include 3 skill-levels
- Hiring lead-time of 1 period (as before)
- No stationary assumptions on turnover, learning and demand are required, **but** for simplicity we assume all three

Estimating Learning and Turnover

- We follow the Maximum Likelihood estimate proposed in Bartholomew, 1991 and use the average past transaction proportions:

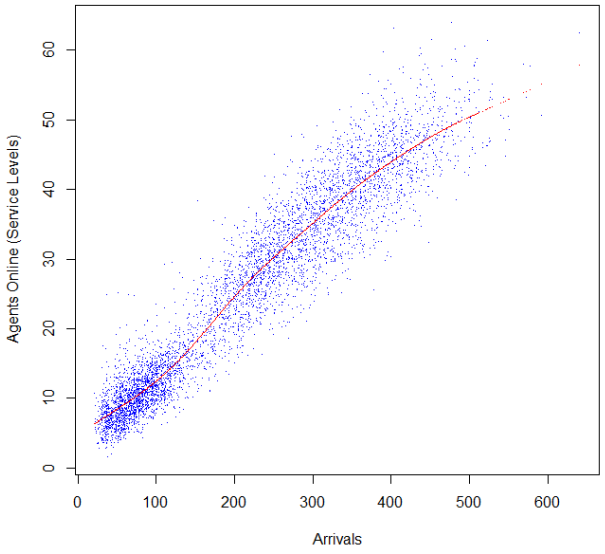
$$l_i = \frac{\bar{n}_{i,i+1}}{\bar{n}_i}$$

- Proportion of learning skill i+1 is estimated with past average proportions of learners:
  - L1 to L2 - 1.5%
  - L2 to L3 - 1.1%
- Total turnover is estimated as in Model 1:

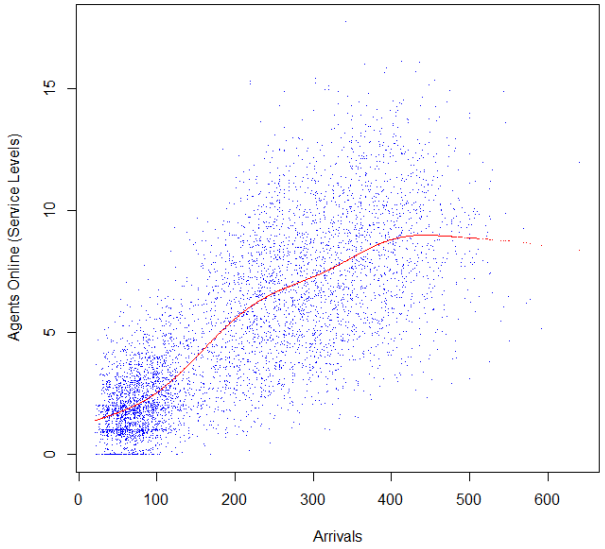
5.27%	Turnover Rate - Stocks	Turnover Rate - Staffing
L1	0.0396	0.0383
L2	0.0084	0.0089
L3	0.0047	0.0055

Half-hour staffing

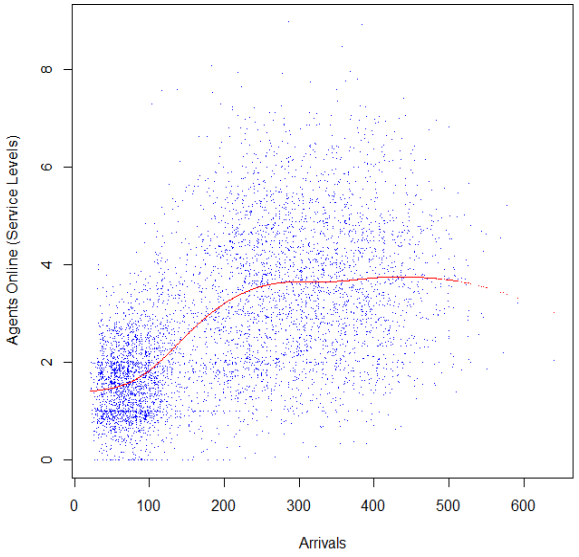
Staffing agents online for all three levels:



Level 1



Level 2



Level 3

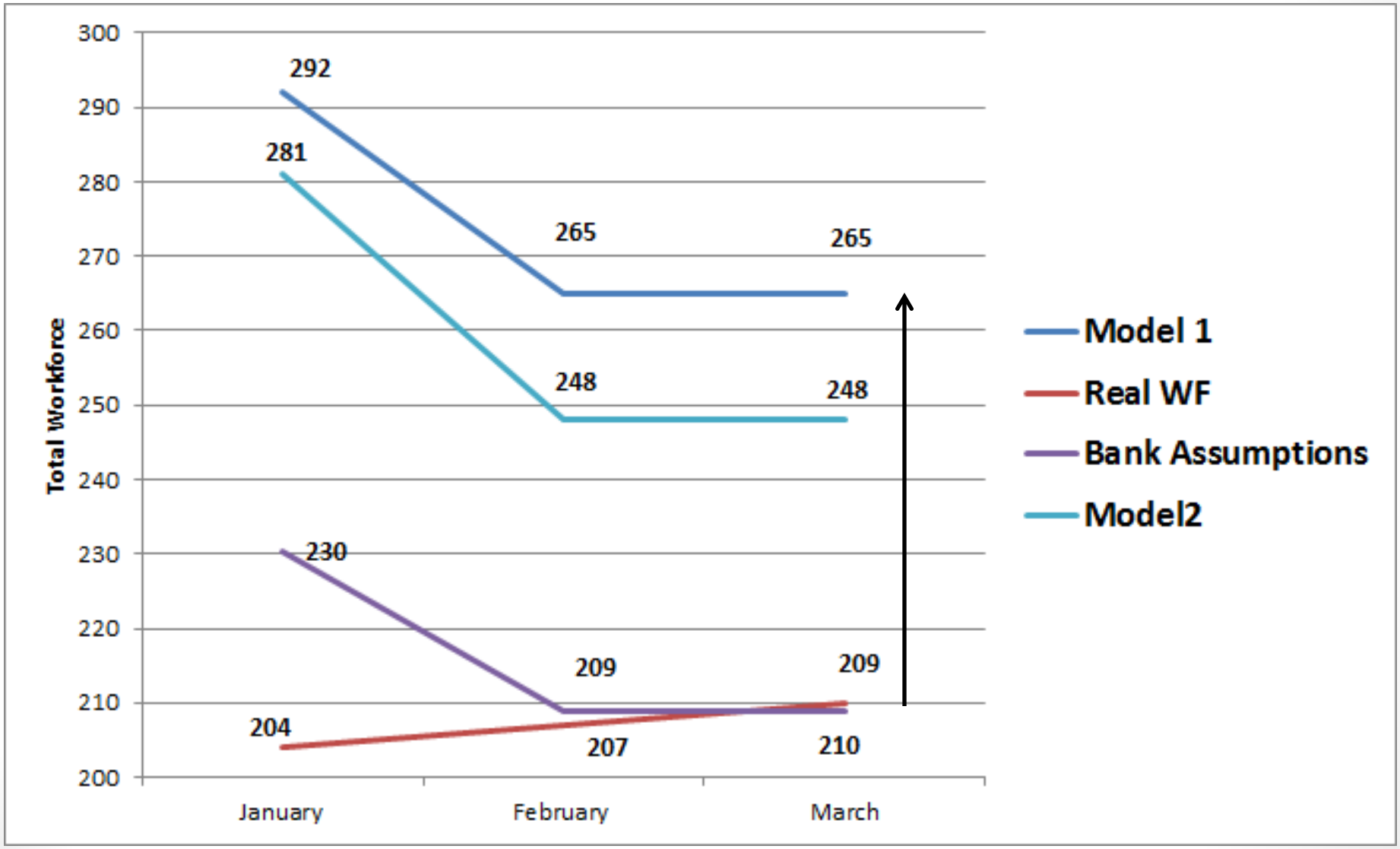
On-job absenteeism is modeled for all three levels
Shift-absenteeism – 12% as before

Model 2: Solution

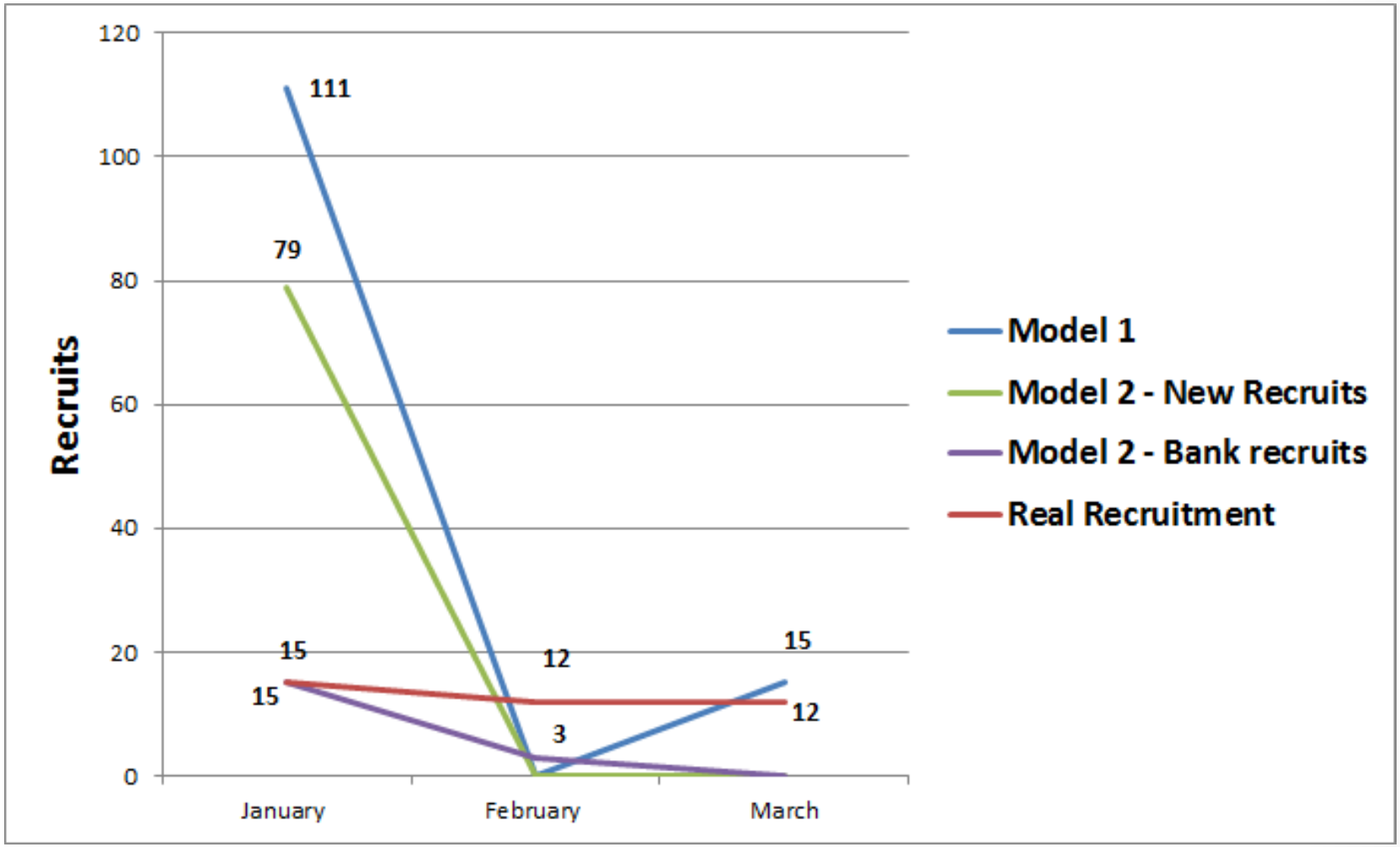
- Due to model assumptions problem is a LP
- **Myopic** "hire-up-to" policy is not necessarily optimal
- Multi-stage "hire-up-to" is promised
- **Problem**: Some skill-levels may be **unattainable** due to low learning proportions
- **Solution**: Bank recruitment (in reality and in our model)

Results Overview

Total Workforce



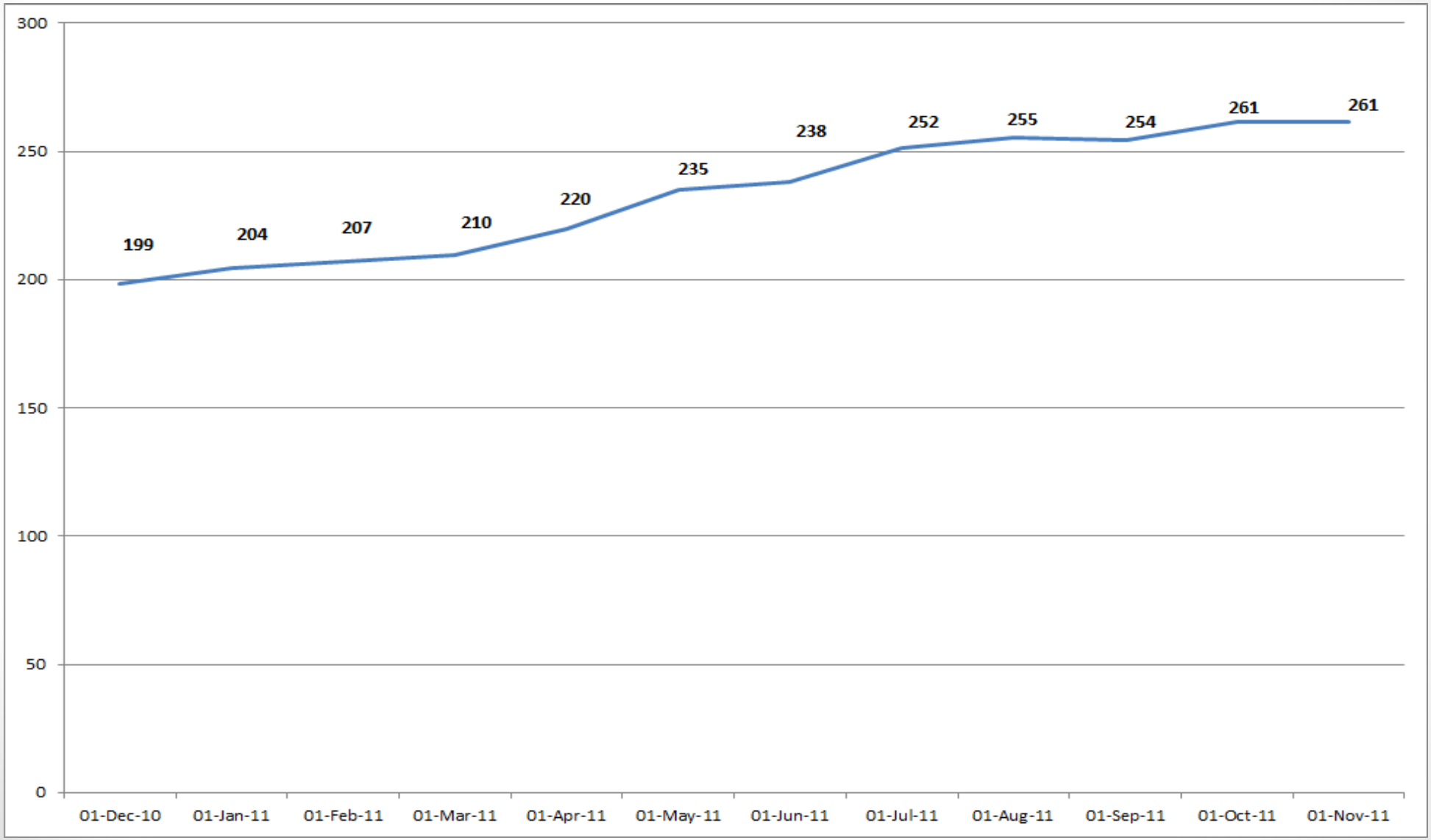
Recruitment



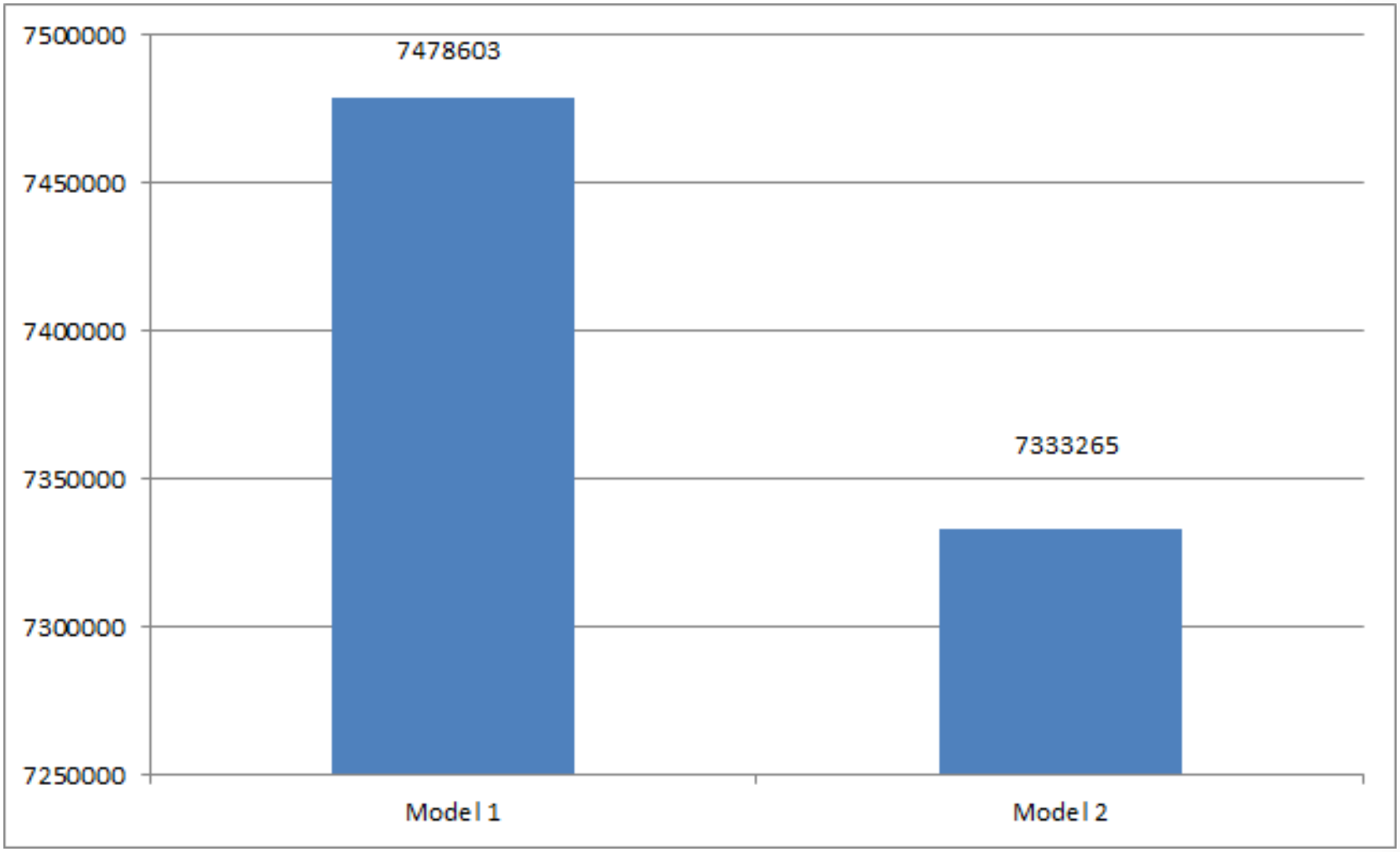
Models vs. Reality

- Uniformly high **service levels** (5%-15% aban. rate)
- **Absenteeism** is accurately estimated (influences peak-hours with high absenteeism rate)
- No overtime assumed – in reality each person is equivalent to more than one full-time employee
- In reality budget "tricks" are possible: Engineer for 3 agents
- Recruitment in large numbers is usually impossible and therefore smoothed
- Having all that said – let us observe reality

In reality – growth is gradual



Comparing Total Costs



Taking learning under consideration can save approx. **153,000 NIS - per quarter**

Why is Model 2 "less expensive"?

- Accurate workforce planning at Level 2 and Level 3
- "Free" recruitment from the bank
- **But**, additional wage is considered for Level 3 employees recruited from bank
- If bank recruitment continues all year then it might be more expensive in the long run (we planned for 1 quarter)
- Bank employees – not infinite pool

Rolling horizon updates

- Planning Horizons are to be selected:
  - Long enough to accommodate Top-Level constraints (recruitment lead-times, turnover,…)
  - Short enough for **stationary assumptions** to hold and **statistical models** to be up to date
    - Improve estimates through newly updated data
- Workforce Planning (cyclical) Algorithm:
  1. Plan a single quarter (or any planning horizon where assumptions hold) using data
  2. Towards the end of planning period update models using new data (demand modeling, staffing function, turnover, learning, absenteeism…)

Future Research

- Solve the full model with the addition of controlled promotion rates
- Prove "hire-up-to" optimality for:
  - Recruitment to all levels (non-linear operating function, stochastic time-varying turnover and learning)
  - Controlled promotions instead of learning
- Validate our models for bank's new data (daily updated)
- Simulation-based optimization for Low-Level planning (Feldman, 2010)
- Server Networks and their applications

Thank you…
Questions/Remarks?