

Big Data, Artificial Intelligence and Data Analysis Set

coordinated by
Jacques Janssen

Volume 9

Data Analysis and Related Applications 1

*Computational, Algorithmic and
Applied Economic Data Analysis*

Edited by
Konstantinos N. Zafeiris
Christos H. Skiadas
Yiannis Dimotikalis
Alex Karagrignoriou
Christiana Karagrignoriou-Vonta

Color Section

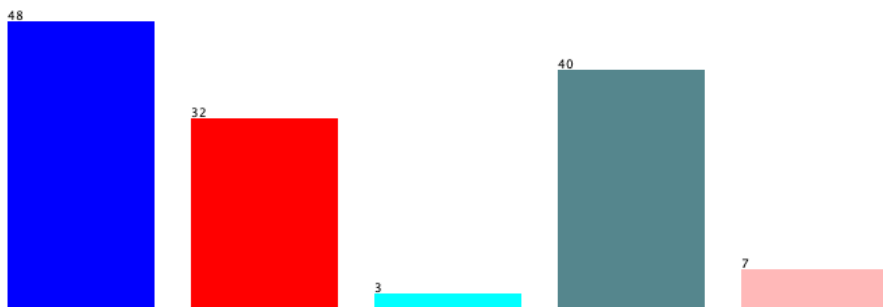


Figure 1.1. *Class visualization for the whole dataset*

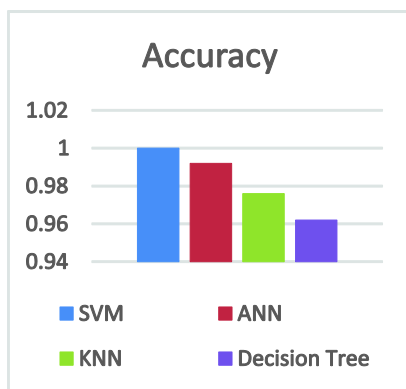


Figure 1.2. *Accuracy comparison*

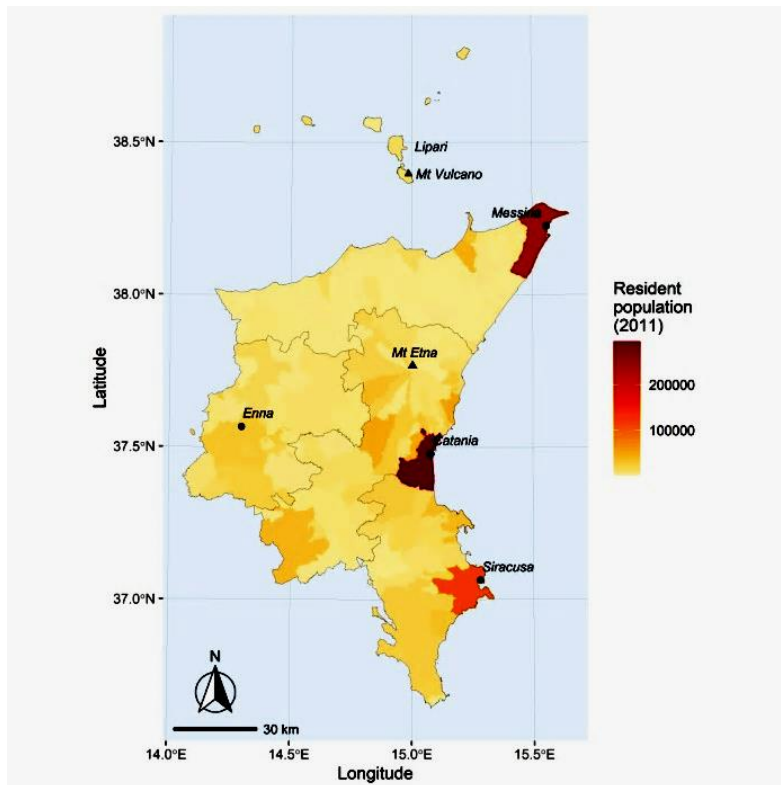


Figure 2.1. Spatial arrangement of resident population.
Source: 15th Italian General Population Census

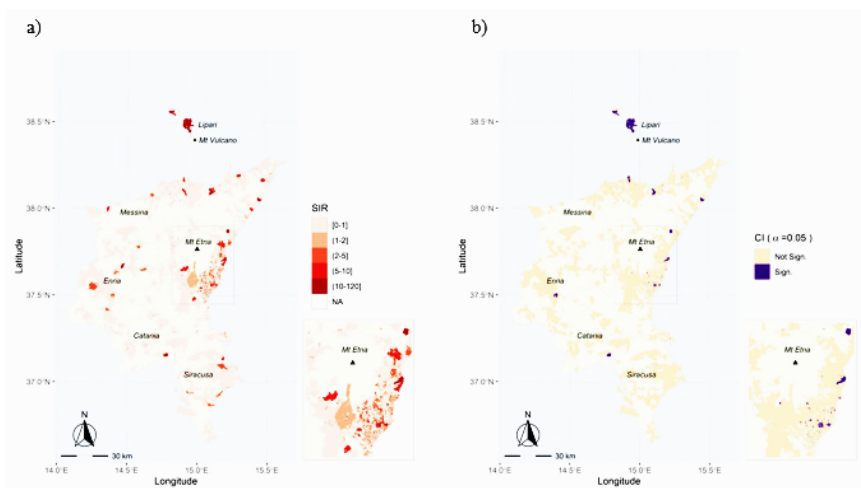


Figure 2.2. SIR distribution by census section (a) and representation of its significance (b). Source: author's elaboration on CRES data

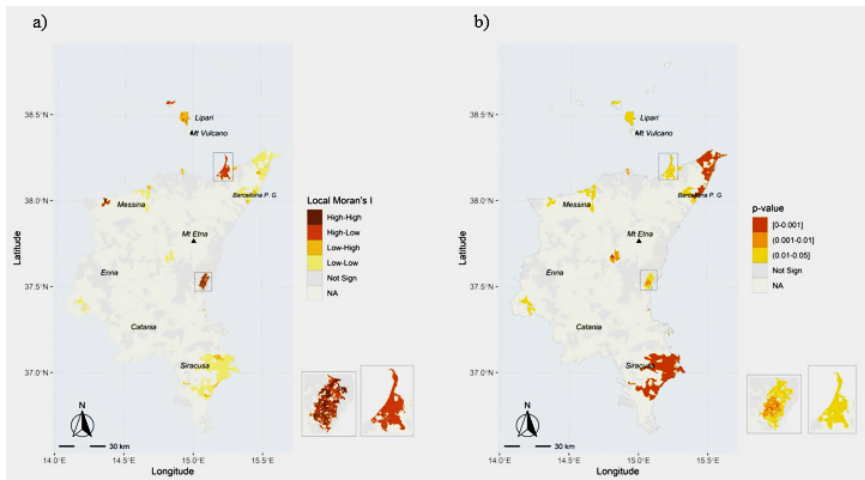


Figure 2.3. Risk cluster map (a) and relative p-values (b). Source: authors' elaboration on CRES data

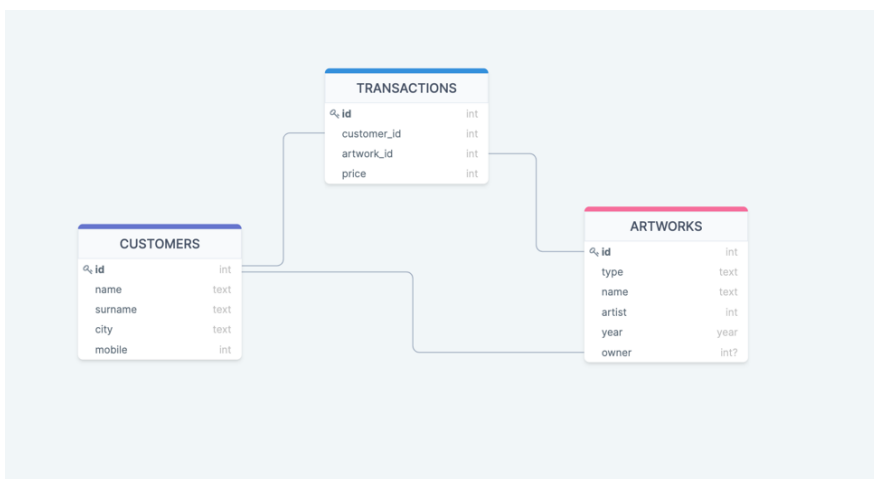


Figure 3.2. Art Shop relational database diagram

```
mysql> LOAD DATA LOCAL INFILE '/root/dummy-data/art-shop-dummy-150.csv' INTO TABLE
artworks FIELDS TERMINATED BY ';' LINES TERMINATED BY '\n' IGNORE 1 ROWS (type,na
me,artist,year,owner);
Query OK, 150 rows affected (0.01 sec)
Records: 150 Deleted: 0 Skipped: 0 Warnings: 0
mysql>
```

Figure 3.4. The process of adding 150 lines of dummy-data with the SQL system and the time elapsed


```

root@nosql-drop:~/nosql# time mongoimport --uri mongodb+srv://admin:adminpass@cluster0.xbf
hm.mongodb.net/nosql-app --collection artworks --type csv --file dummy-data/art-shop-dummy
-150.csv --headerline
2021-07-02T10:05:21.807+0000    connected to: localhost
2021-07-02T10:05:21.835+0000    imported 150 documents

real    0m0.208s
user    0m0.027s
sys     0m0.015s

```

Figure 3.5. *The process of adding 150 lines of dummy-data with the NoSQL system and the time elapsed*

```

gitpod /workspace/starport/artshop $ time artshopd tx artshop create-artwork 'Painting' 'Name2' 'Artist 2' '1999' '0' --from=dbtests
{"body":{"messages":[{"type":"/bozonet.artshop.artshop.MsgCreateArtwork","creator":"cosmosiu04hd2m8laexpj0msajps8smt9yf9d9kx8cvv8","
Artytype":"Painting","name":"Name2","artist":"Artist 2","year":"1999","owner":"0"},"memo":"","timeout_height":"0","extension_options"
:[],"non_critical_extension_options":[]},"auth_info":{"signer_infos":[]},"fee":{"amount":[],"gas_limit":"200000","payer":"","granter":
""},"signatures":[]}}
confirm transaction before signing and broadcasting [y/N]: y
{"height":"331","txhash":"A39468C88C03385918C55B64550BE66833135DFCB7C945E5E3A8EC84D292BF","codespace":"","code":0,"data":"0A130A0D4
37265617465417274776F726812028081","raw_log":[{"type":"message","attributes":[{"key":"action","value":"CreateArtwork"}]}]}","log":["msg_index:0","log":"","events":[{"type":"message","attributes":[{"key":"action","value":"CreateArtwork"}]}]}],"info":"","gas_wanted":"200000","gas_used":"45824","tx":null,"timestamp":""}

real    0m1.807s
user    0m0.071s
sys     0m0.032s

```

Figure 3.6. *Adding a single entry to the blockchain-supported system and the time elapsed*

```

! config.yml > ...
1  accounts:
2  |   - name: dbtests
3  |     coins: ["20000token", "200000000stake"]
4  |   - name: bob
5  |     coins: ["10000token", "100000000stake"]
6  validator:
7  |   name: dbtests
8  |   staked: "100000000stake"
9  client:
10 |   vuex:
11 |     path: "vue/src/store"
12 |   openapi:
13 |     path: "docs/static/openapi.yml"
14 faucet:
15 |   name: bob
16 |   coins: ["5token", "100000stake"]
17

```

Figure 3.8. *Starport config.yml file*



Figure 5.1. 25 de Abril Bridge and the Sanctuary of Christ the King monument (to the right of the photo) in the city of Almada. The photo was taken by the first author in September 2019

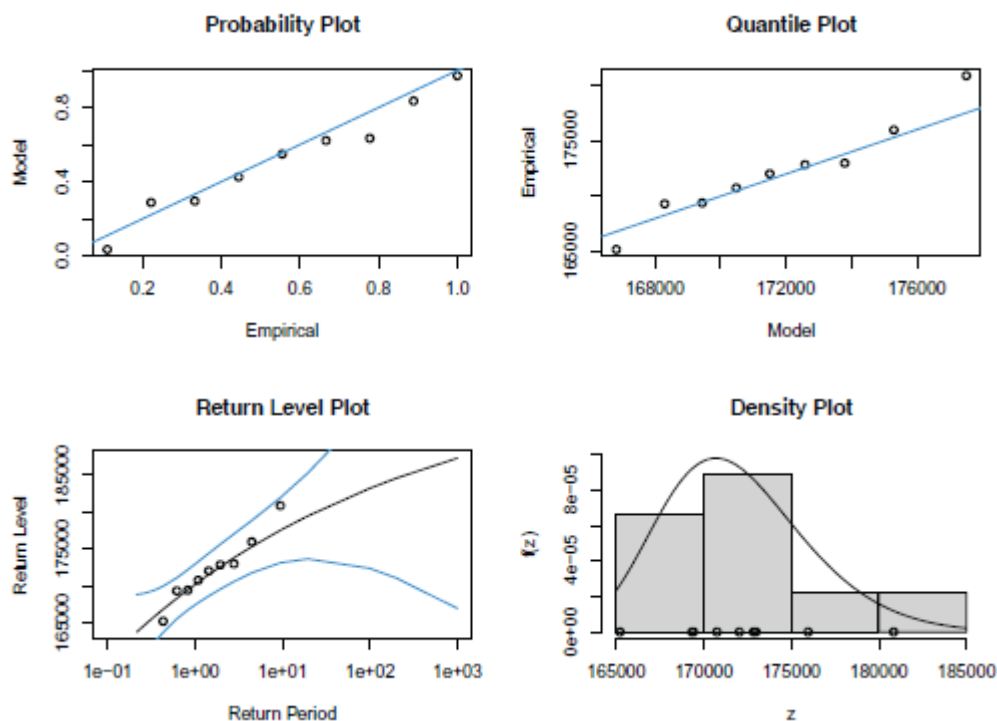


Figure 5.4. Diagnostic plots of the GEV model fit to the yearly maximum from the daily traffic data of the 25 de Abril Bridge

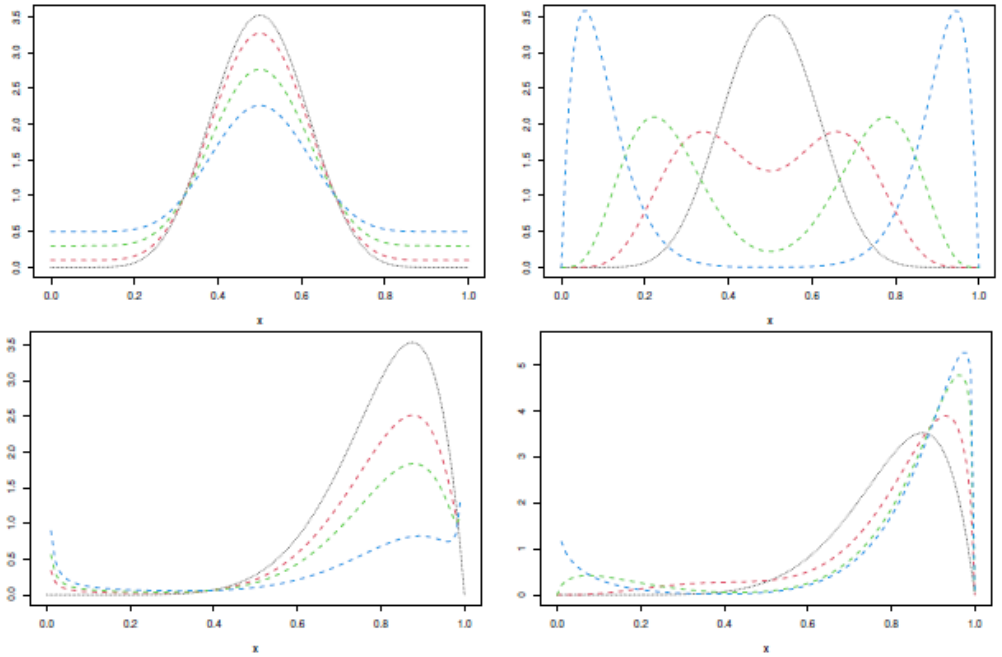


Figure 8.1. Top panels: the dotted line refers to the density of a $\text{Beta}(0.5, 20)$. On the left-hand side, the dashed lines refer to densities from a VIB rv with $\mu = 0.5$, $\varphi = 20$, $k = 0.1$ and $p = \{0.1$ (red), 0.3 (green), 0.5 (blue)}. On the right-hand side, the dashed lines refer to densities from an FB rv with $\mu = 0.5$, $\varphi = 20$, $p = 0.5$ and $w = \{0.3$ (red), 0.5 (green), 0.8 (blue)}. Bottom panels: the dotted line refers to the density of a $\text{Beta}(0.8, 10)$. On the left-hand side, the dashed lines refer to densities from a VIB rv with $\mu = 0.8$, $\varphi = 10$, $k = 0.01$ and $p = \{0.3$ (red), 0.5 (green), 0.8 (blue)}. On the right-hand side, the dashed lines refer to densities from an FB rv with $\mu = 0.8$, $\varphi = 10$, $p = 0.9$ and $w = \{0.5$ (red), 0.8 (green), 0.9 (blue)}

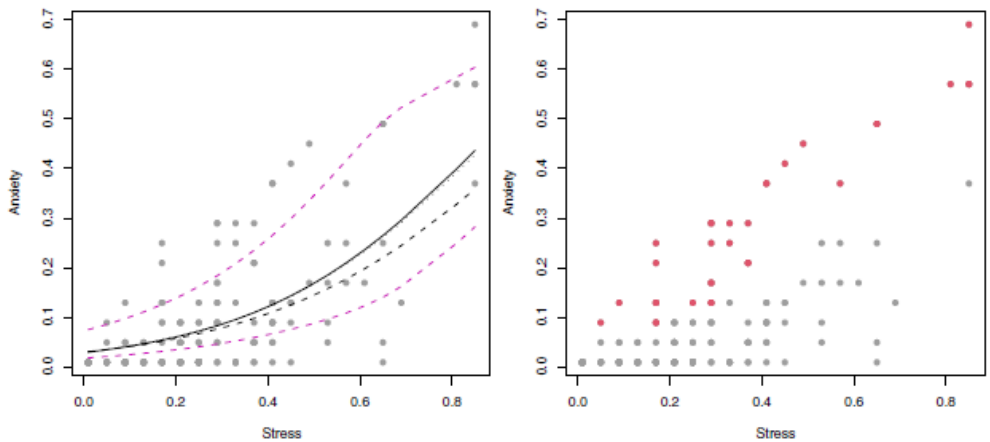


Figure 8.2. Left-hand side: fitted regression curves for the models. Breg in dotted line, VIBreg in solid line and FBreg in dashed lines. Colored dashed curves refer to the component means of the FBreg model. Right-hand side: scatterplot of stress level versus anxiety level. Red dots refer to subjects belonging to group 1

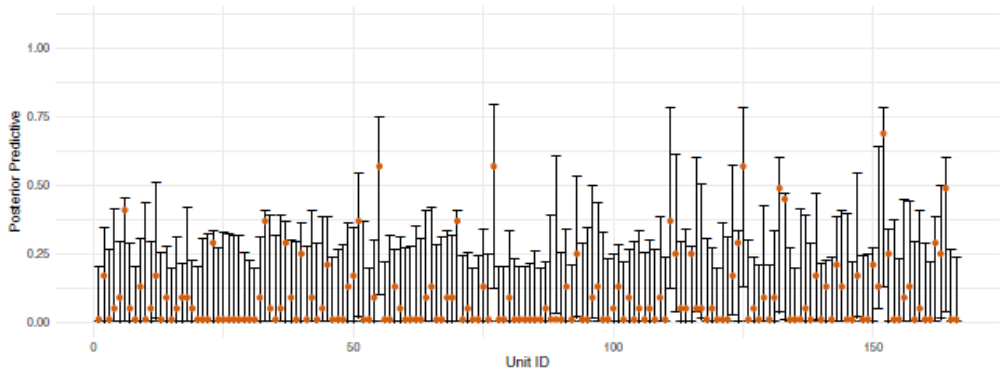


Figure 8.3. 95% posterior predictive intervals for each statistical unit for the VIBreg model. The observed anxiety levels are represented with orange dots

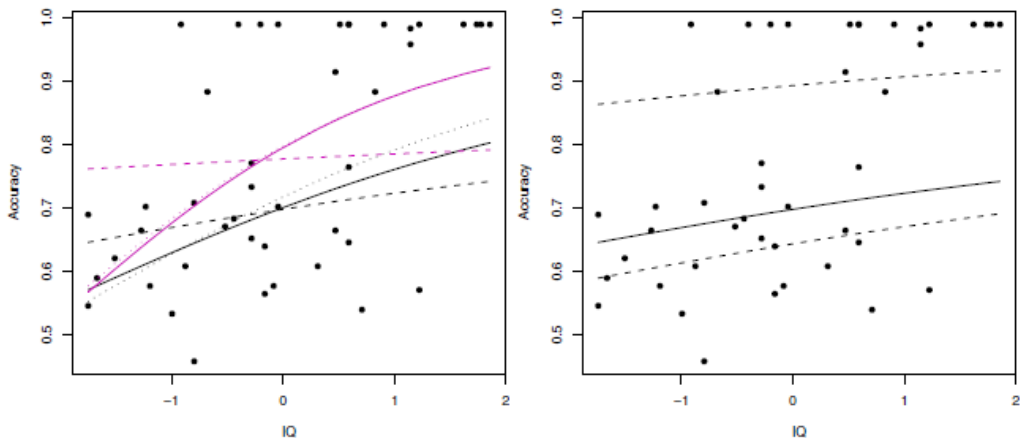


Figure 8.4. Left-hand side: fitted regression curves for the models with augmentation (violet lines) and without augmentation (black lines) refer to the Breg (dotted lines), VIBreg (solid lines) and FBreg (dashed lines) models. Right-hand side: fitted regression curve for the overallmean (solid line) and for the componentmeans (dashed lines) of the AFBreg model

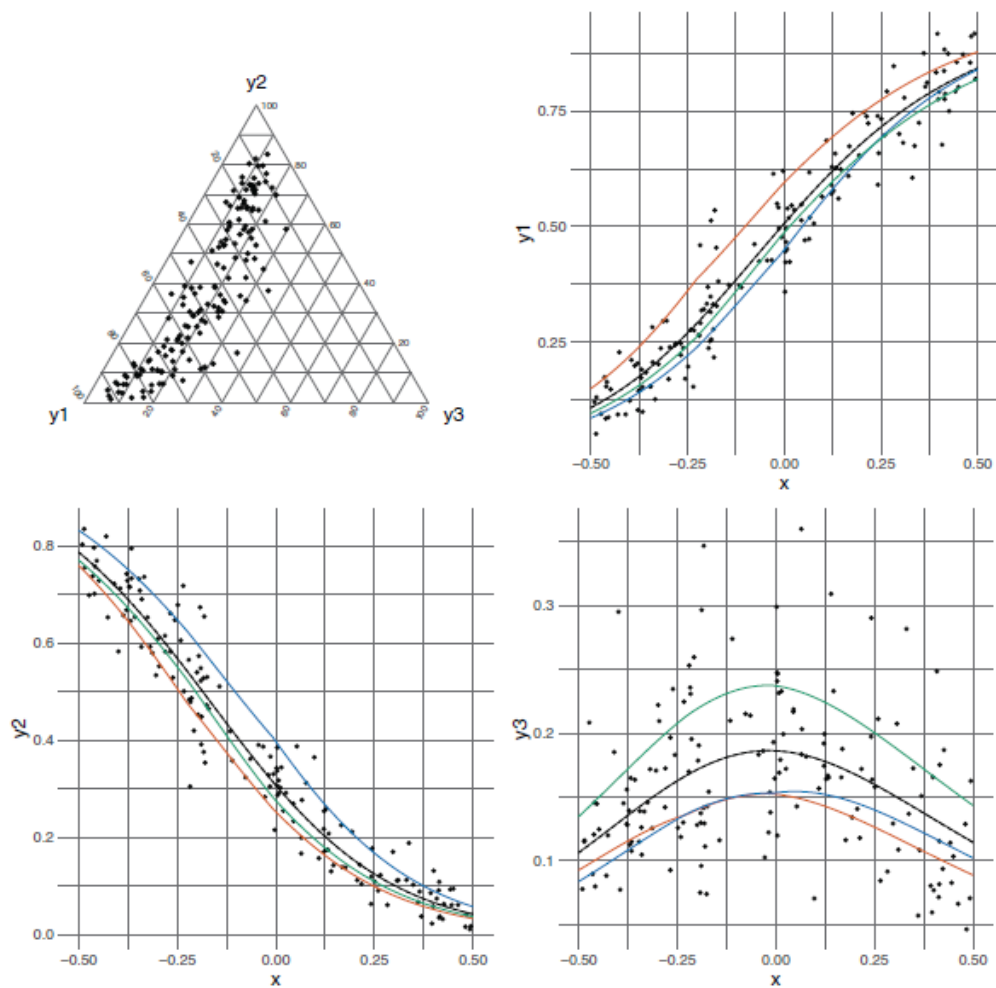


Figure 9.1. Representations of one replication from scenario (i)

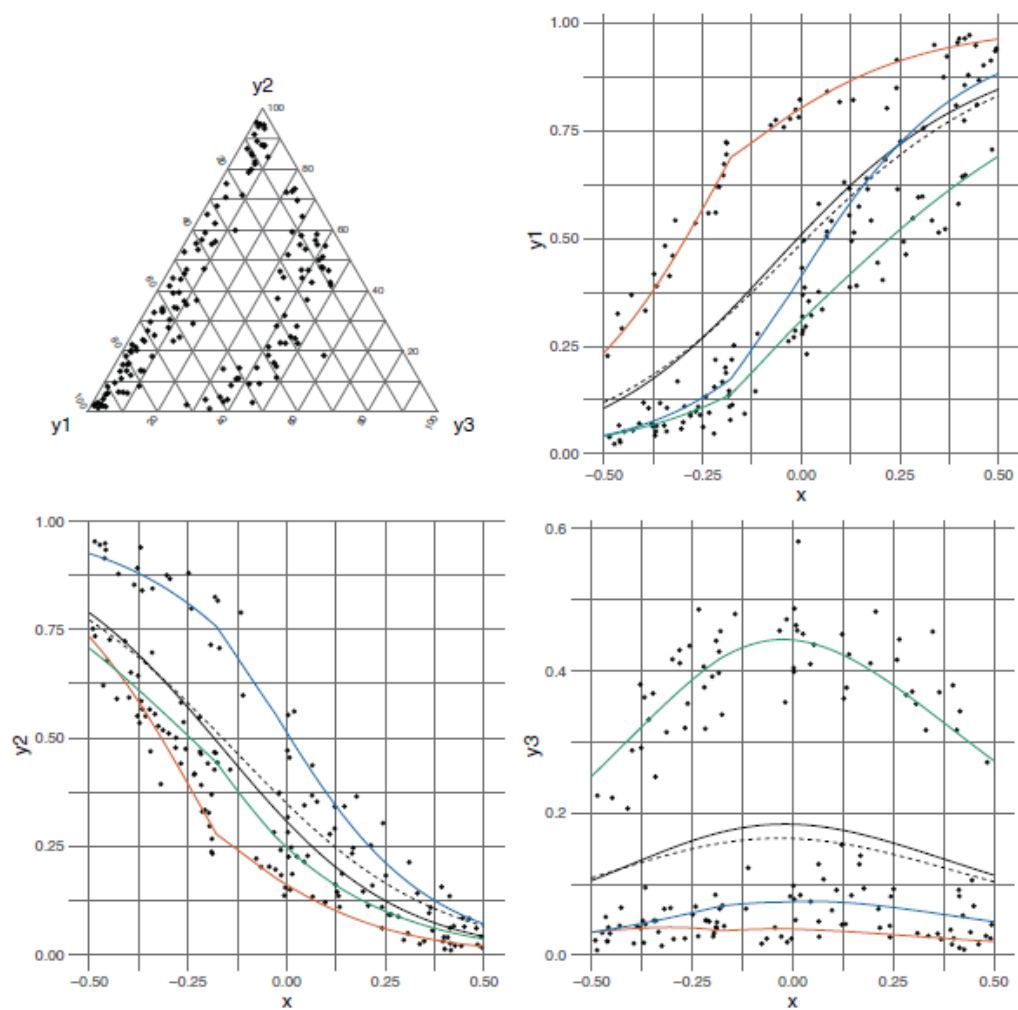


Figure 9.2. Representations of one replication from scenario (ii)

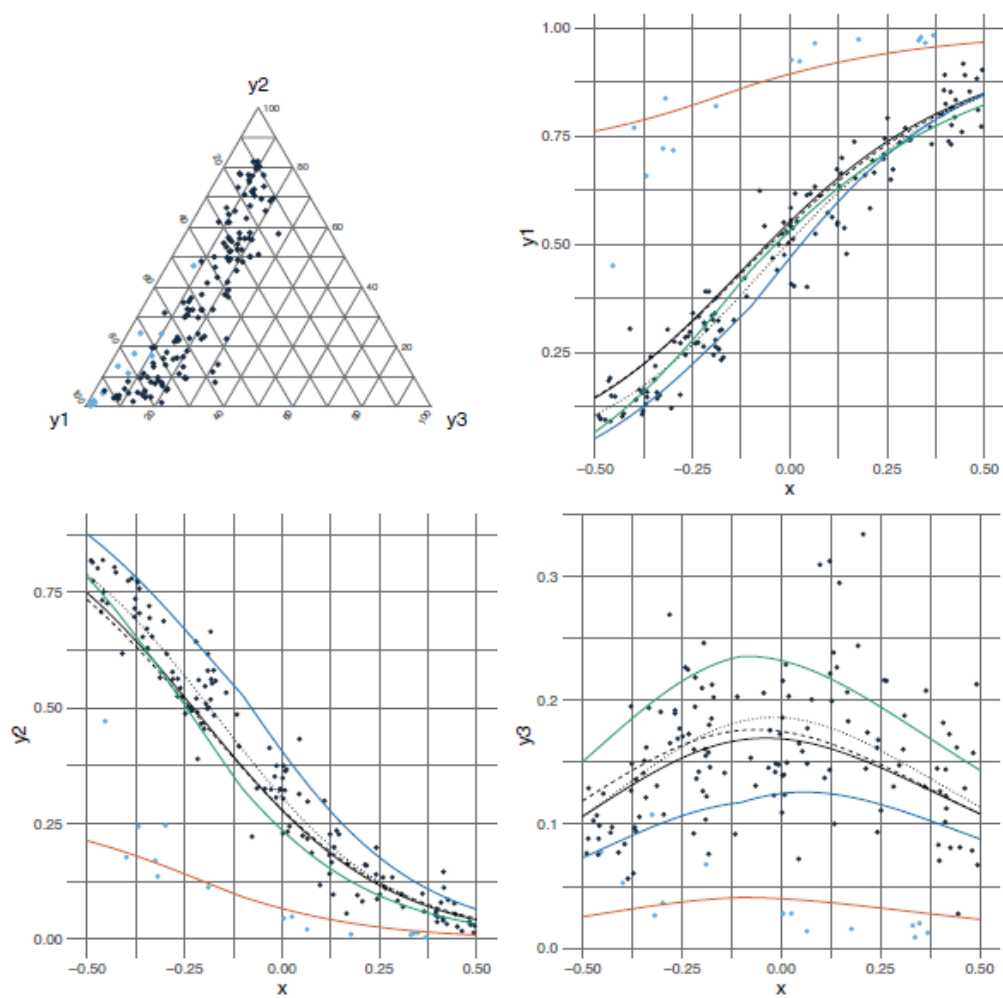


Figure 9.3. Scenario (I). Perturbed points are in light-blue and unperturbed points are in black

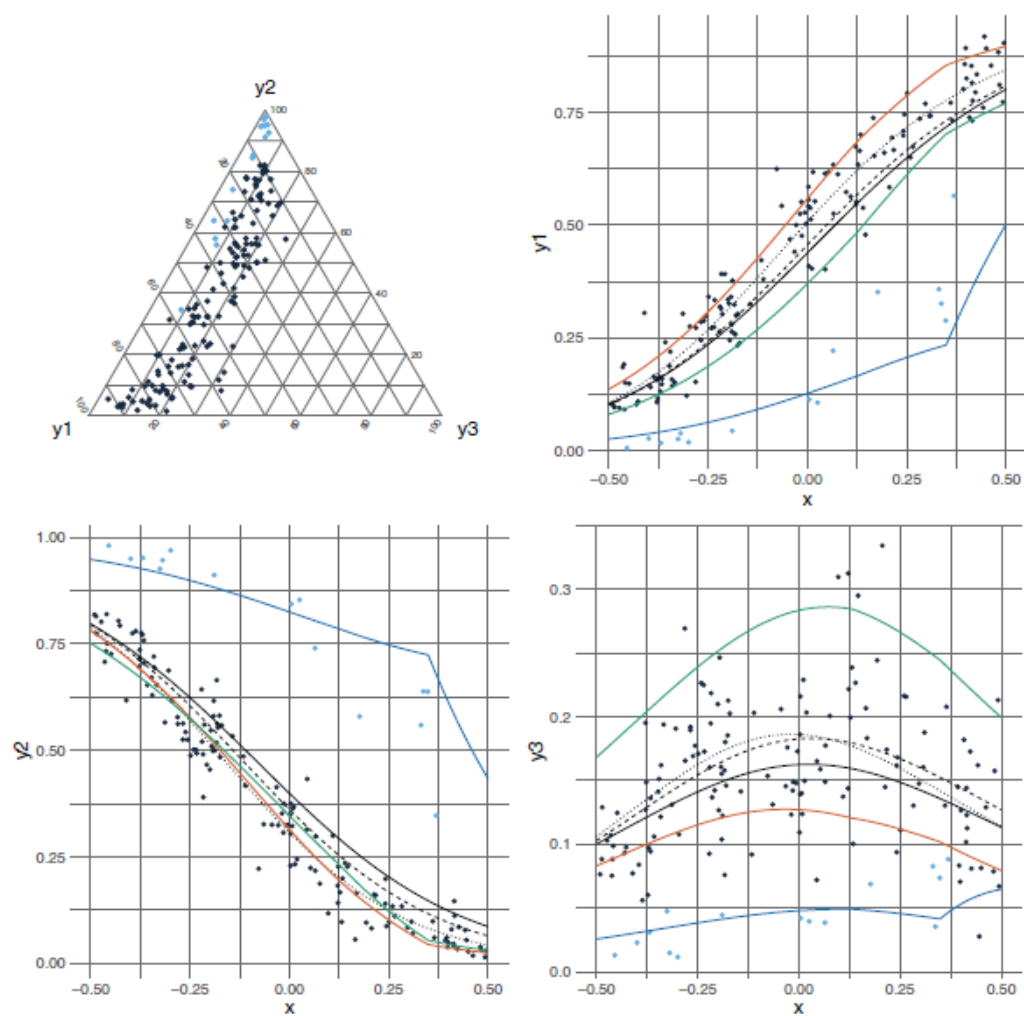


Figure 9.4. Scenario (II). Perturbed points are in light-blue and unperturbed points are in black

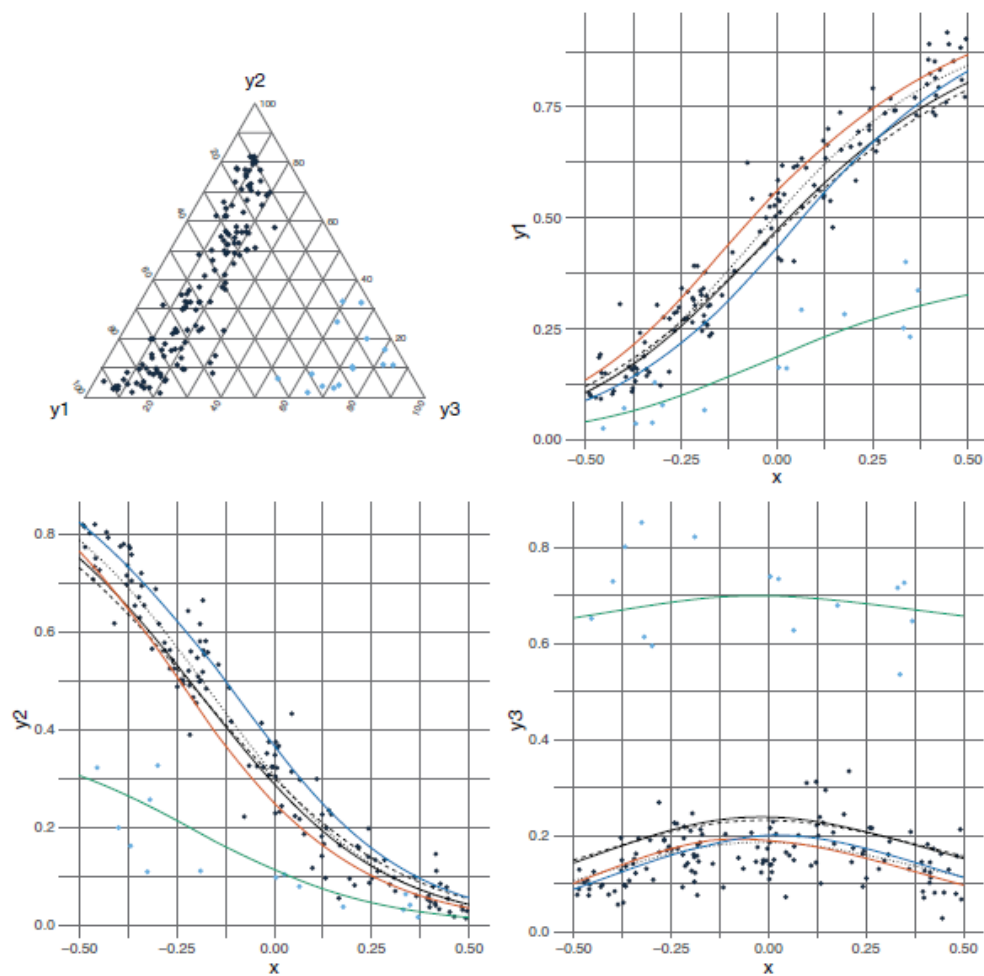


Figure 9.5. Scenario (III). Perturbed points are in light-blue and unperturbed points are in black

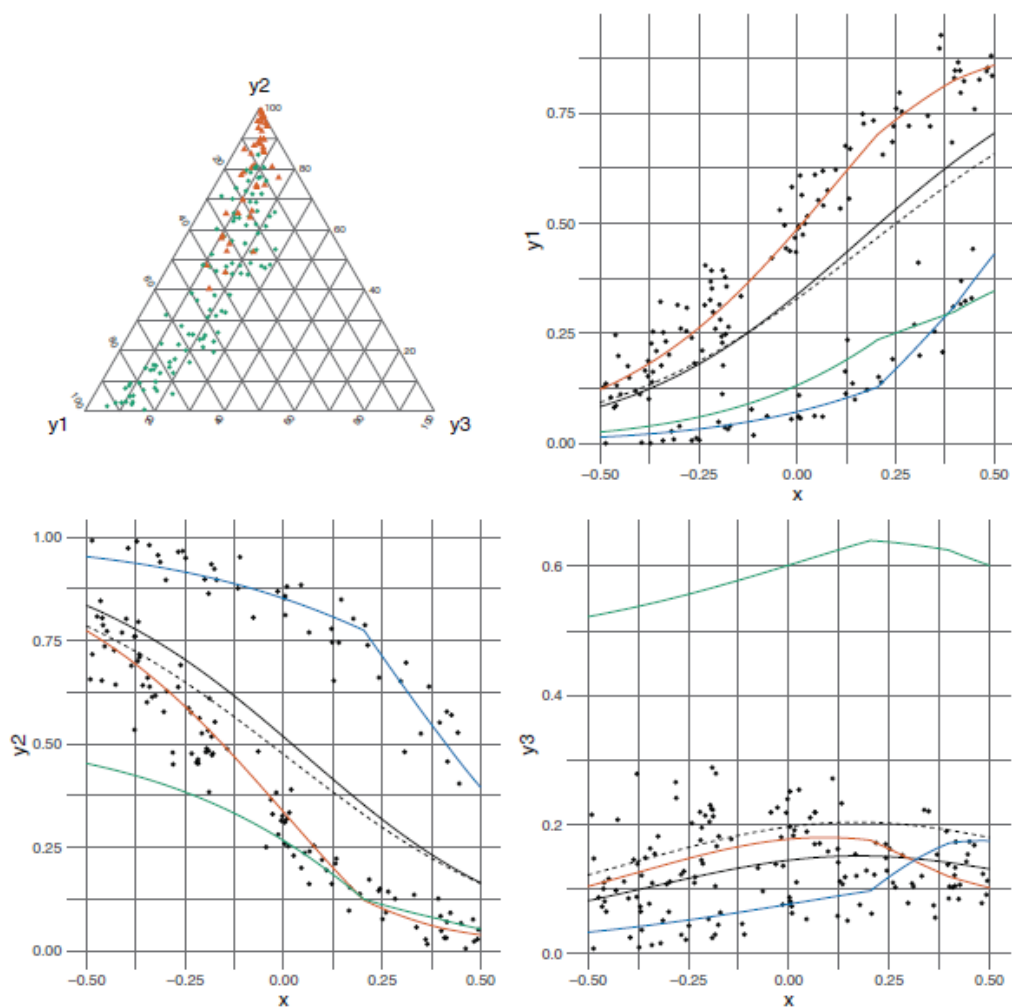


Figure 9.6. Representations of one replication in the presence of latent groups, scenario (a)

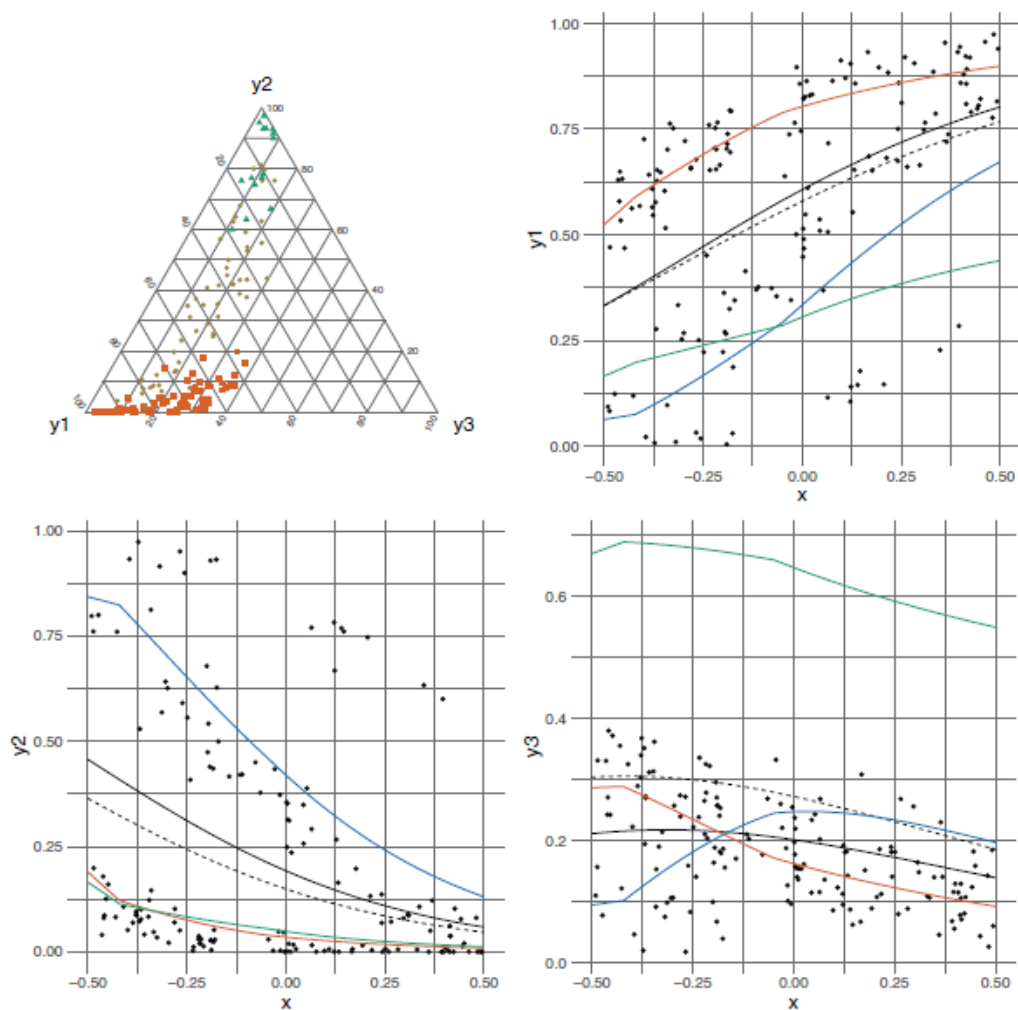


Figure 9.7. Representations of one replication
in the presence of latent groups, scenario (b)

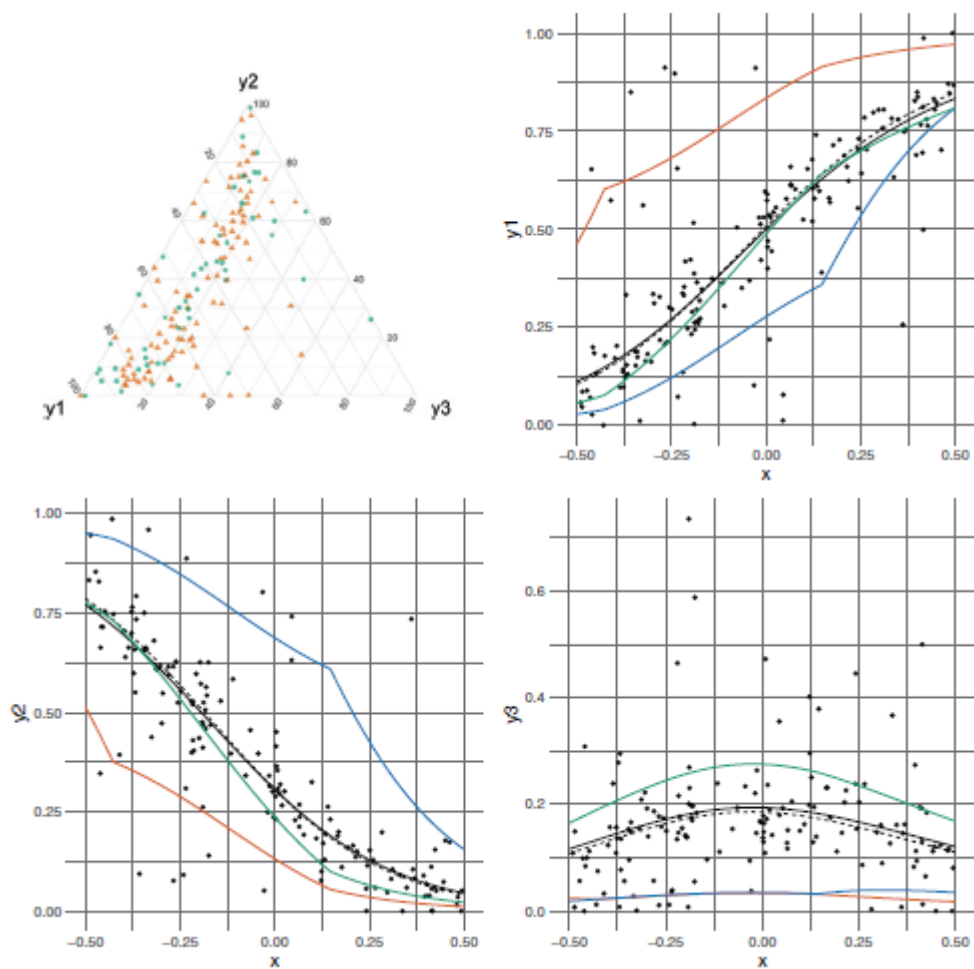


Figure 9.8. Representations of one replication of a generic mixture of two Dirichlet distributions

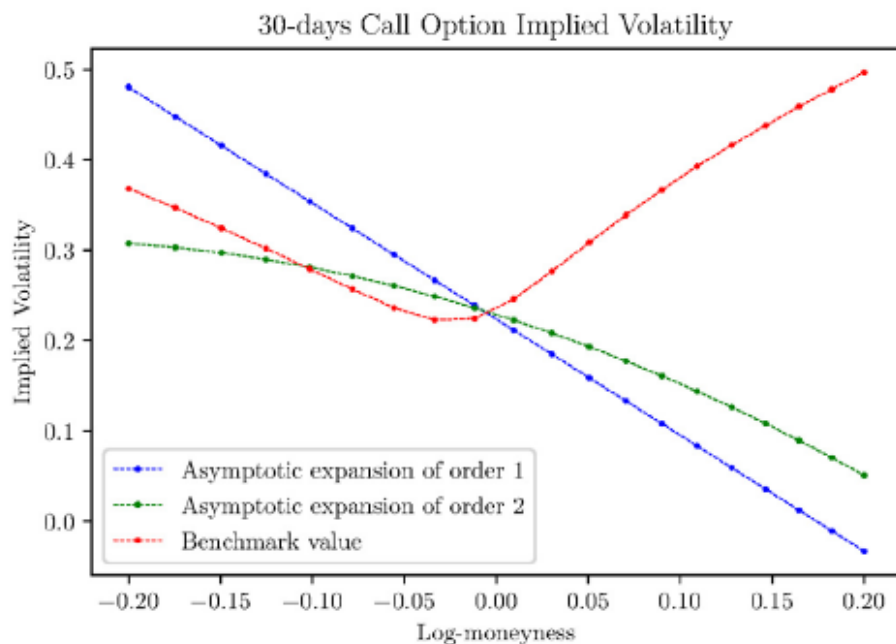


Figure 10.1. An example of the asymptotic expansions of orders 1 and 2 of the implied volatility, and the benchmark value for a call option with a maturity of 30 days

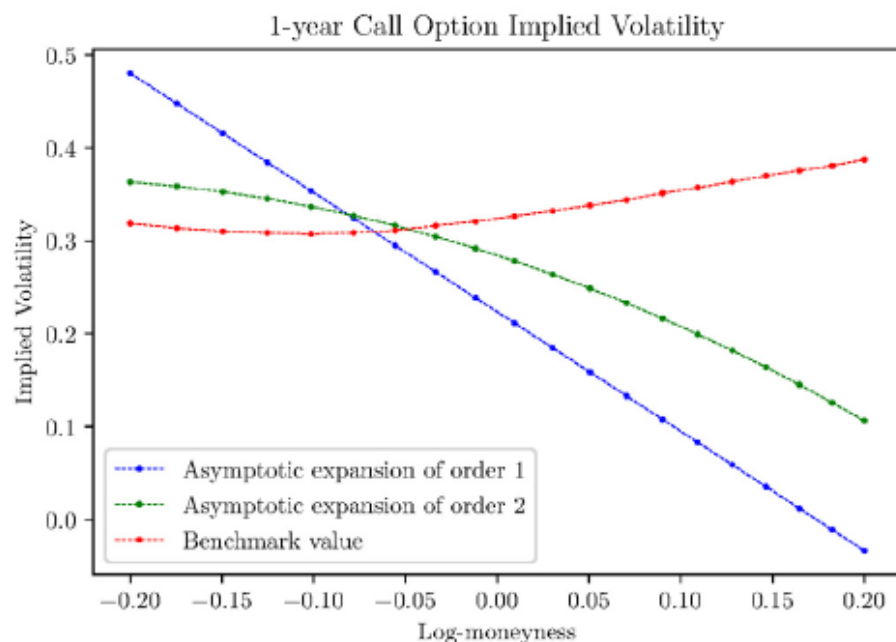


Figure 10.2. An example of the asymptotic expansions of orders 1 and 2 of the implied volatility, and the benchmark value for a call option with a maturity of 1 year

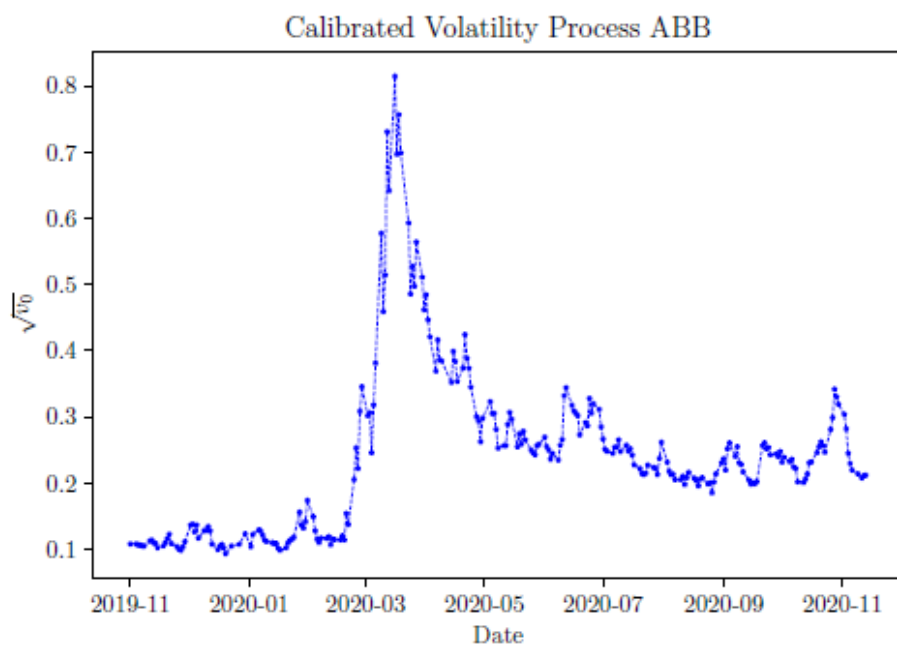


Figure 10.3. Calibrated daily values for the daily volatility $\sqrt{v_0}$ using ABB stock data

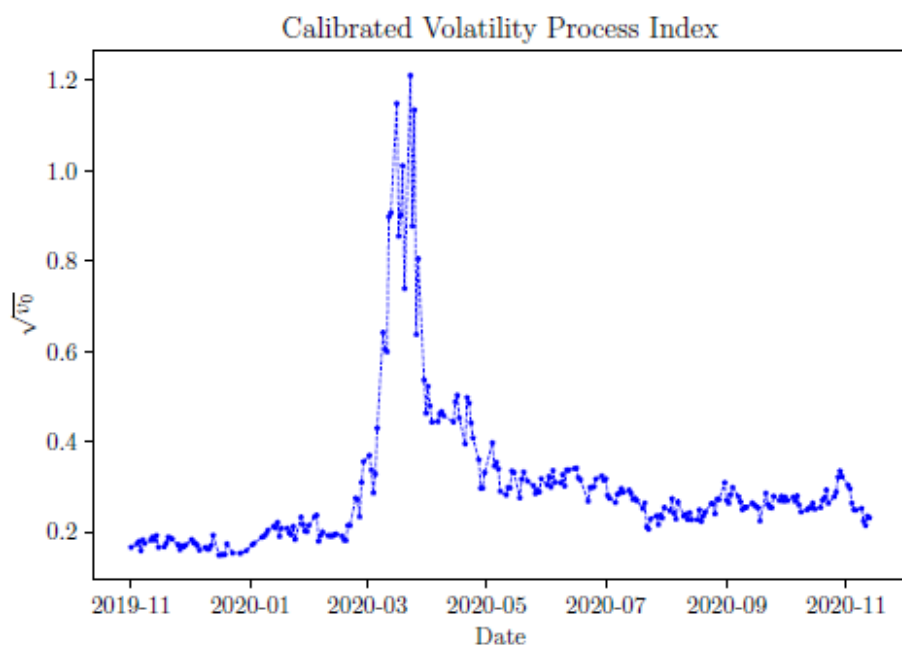


Figure 10.4. Calibrated daily values for the daily volatility $\sqrt{v_0}$ using Eurostock 50 Index data

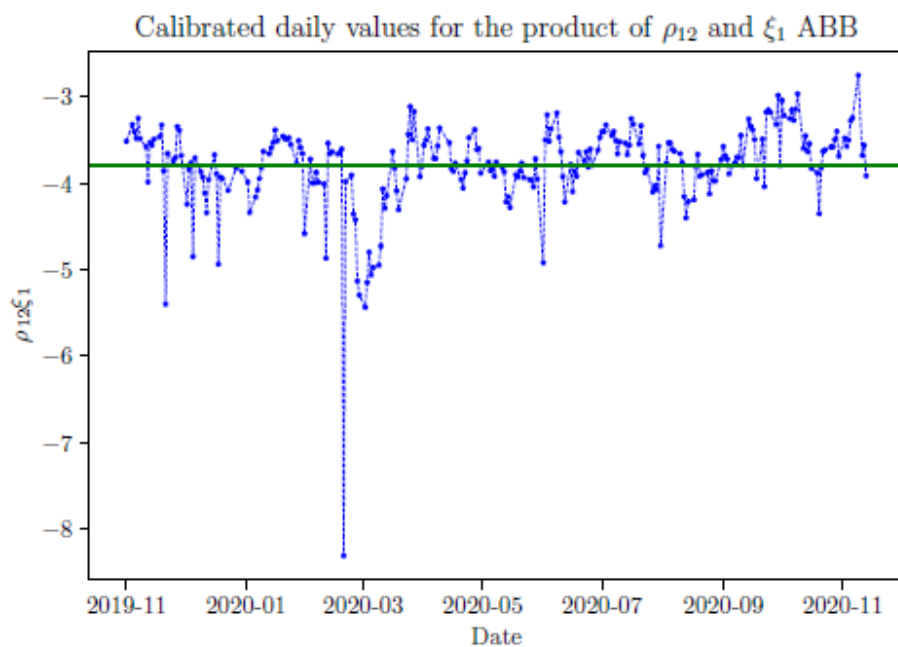


Figure 10.5. Calibrated daily values for the product of ρ_{12} and ξ_1 , $\rho_{12}\xi_1$, using ABB stock data

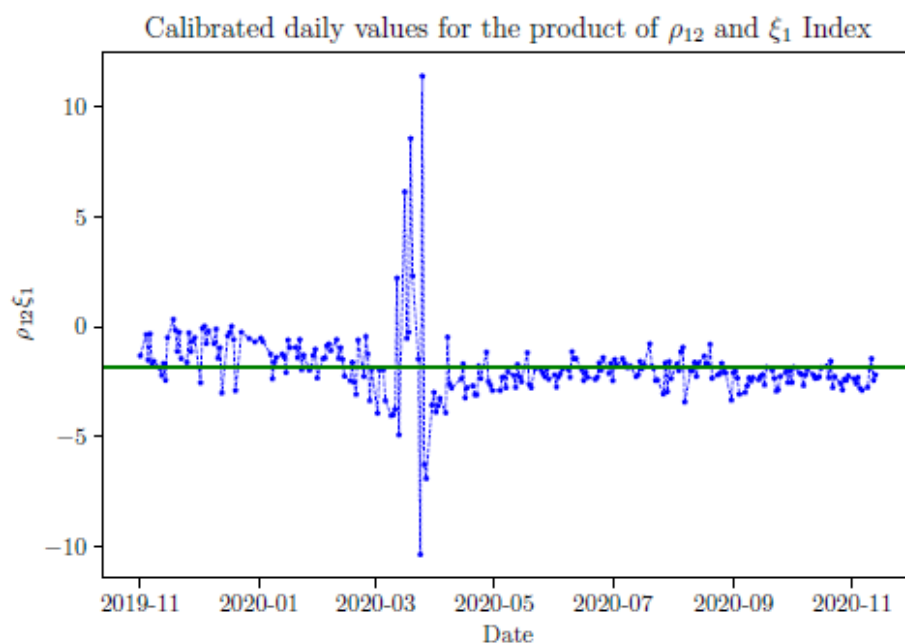


Figure 10.6. Calibrated daily values for the product of ρ_{12} and ξ_1 , $\rho_{12}\xi_1$, using Eurostock 50 Index data

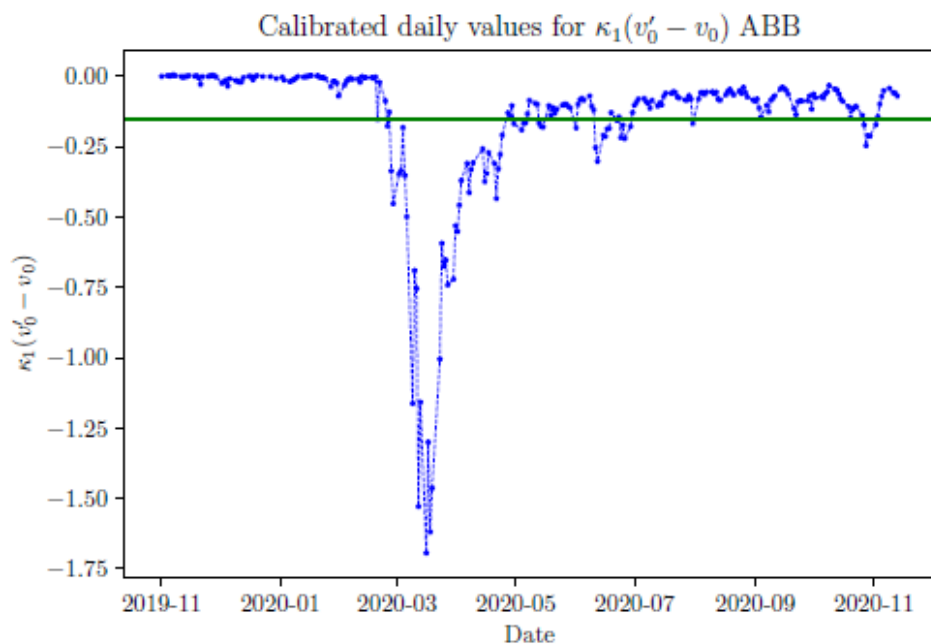


Figure 10.7. Calibrated daily values for the product $\kappa_1(v'_0 - v_0)$ using ABB stock data

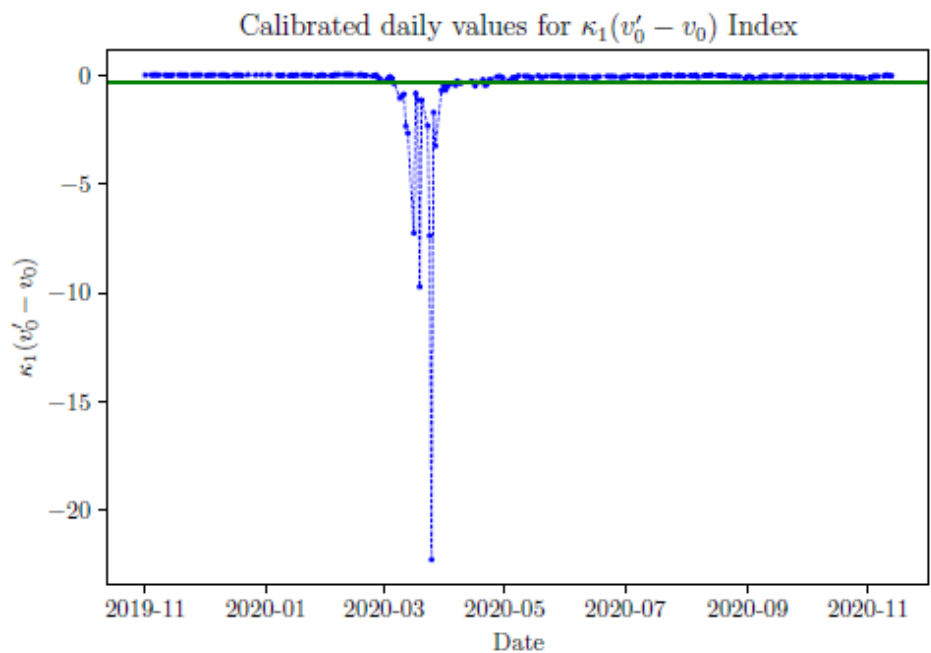


Figure 10.8. Calibrated daily values for the product $\kappa_1(v'_0 - v_0)$ using Eurostock 50 Index data

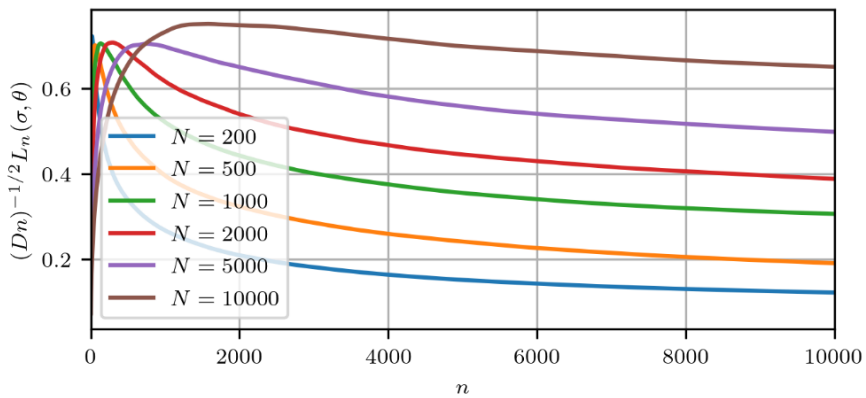


Figure 12.1. Maximum scaled regret versus step number for $c_2 - c_1 = 10$

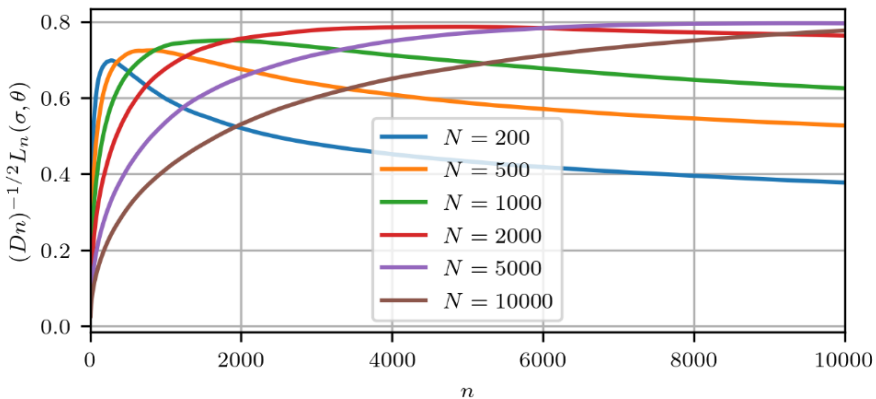


Figure 12.2. Maximum scaled regret versus step number $c_2 - c_1 = 3.3$

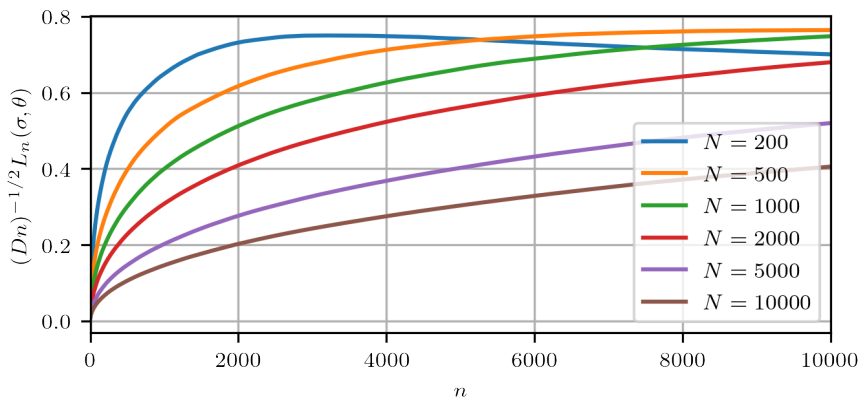


Figure 12.3. Maximum scaled regret versus step number $c_2 - c_1 = 1$

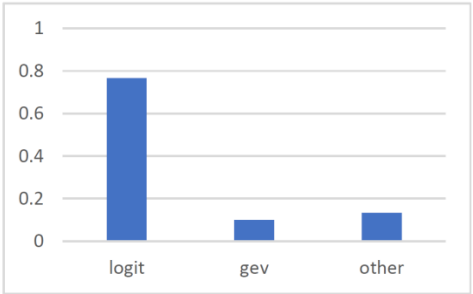


Figure 13.1. *Percentage of choice (logit dataset)*

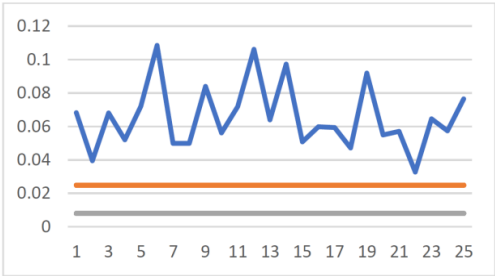


Figure 13.2. *Comparison between AIC and BIC, blue = $\Delta\log(\text{MSE})$, gray = 0.008, orange = 0.024858*

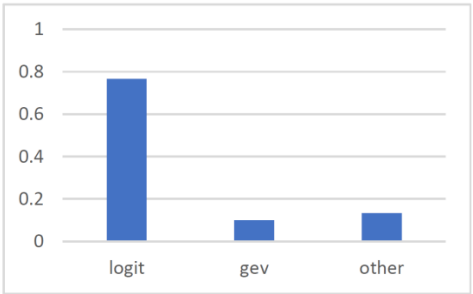


Figure 13.1. *Percentage of choice (logit dataset)*

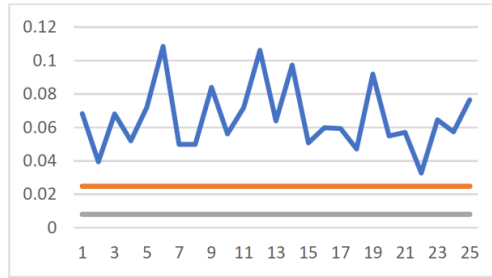


Figure 13.2. Comparison between AIC and BIC, blue = $\Delta \log(\text{MSE})$, gray = 0.008, orange = 0.024858

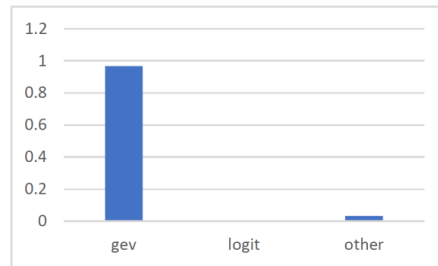


Figure 13.5. Percentage of choice (gev data)

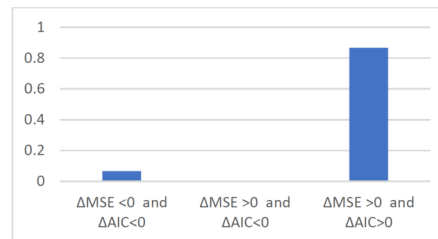


Figure 13.6. Comparison of AIC and of MSE (if $\Delta \text{MSE} > 0$, according to MSE the best model is logev)

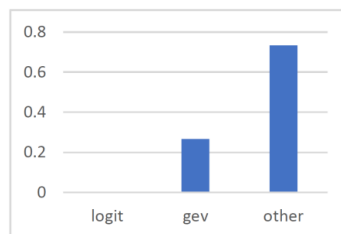


Figure 13.7. Percentage of choice (non-monotonic data)

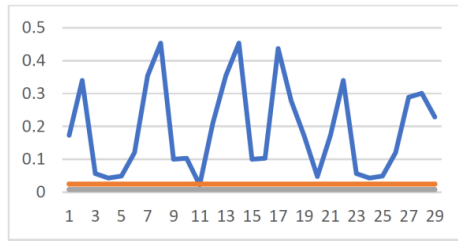


Figure 13.8. Comparison between AIC and BIC, blue = $\Delta \log(\text{MSE})$, gray = 0.008 and orange = 0.024858

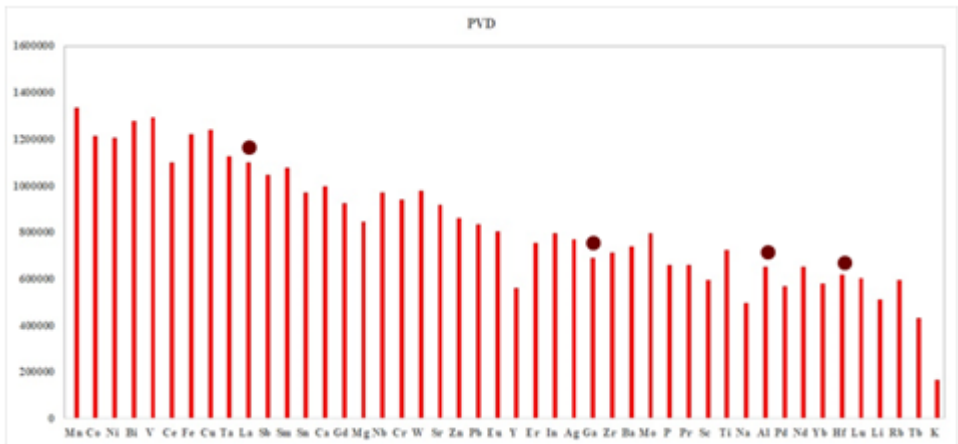


Figure 14.3(a). Dataset derived from the physical vapor deposition device for various high- k dielectric and semiconductor materials

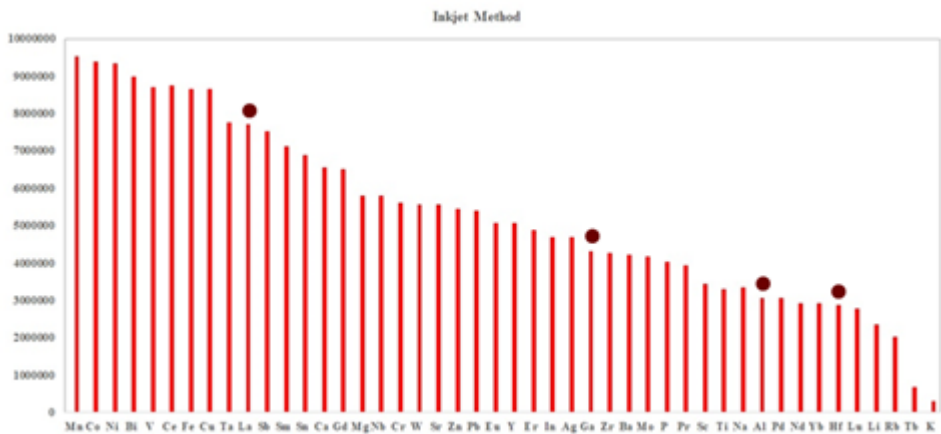


Figure 14.3(b). Dataset derived from the inkjet method for various high- k dielectric and semiconductor materials

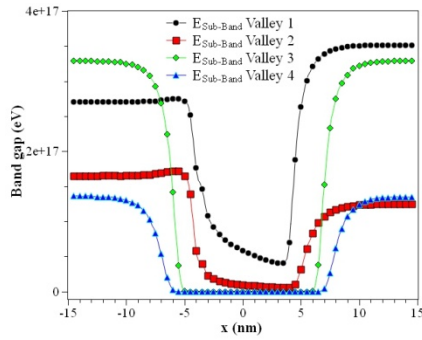


Figure 14.4. Simulation data for energy sub-bands for valleys of the DG MOSFET

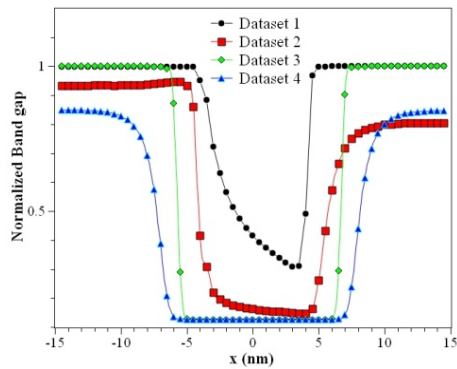


Figure 14.5. Normalized dataset for energy sub-bands attained for the same

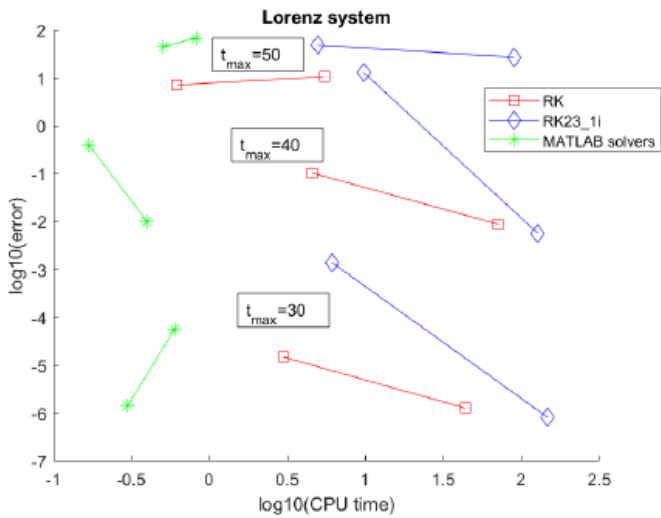


Figure 15.1. Efficiency comparison (error in the $\| \cdot \|_1$ -norm)

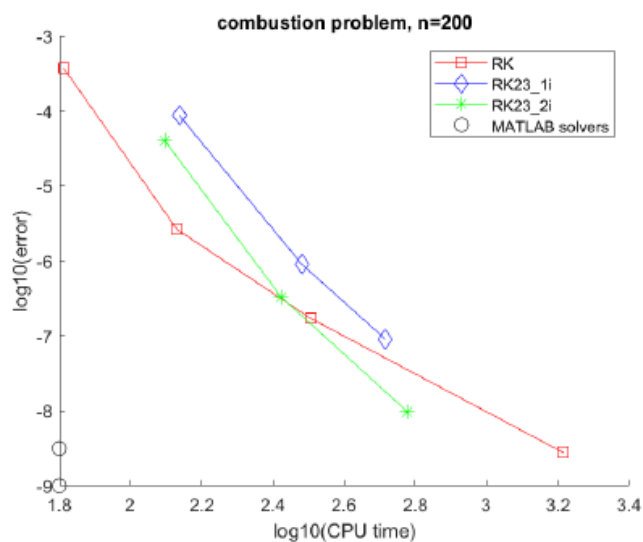


Figure 15.2. Efficiency comparison (error in the max-norm) for $n = 200$

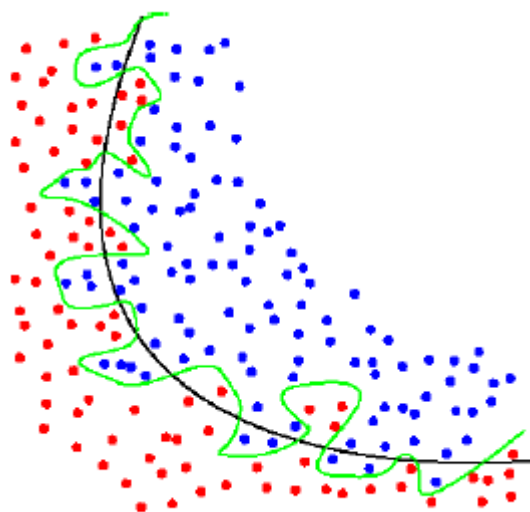


Figure 16.1. Example decision boundaries



Figure 16.2. *Sampling of a circle*

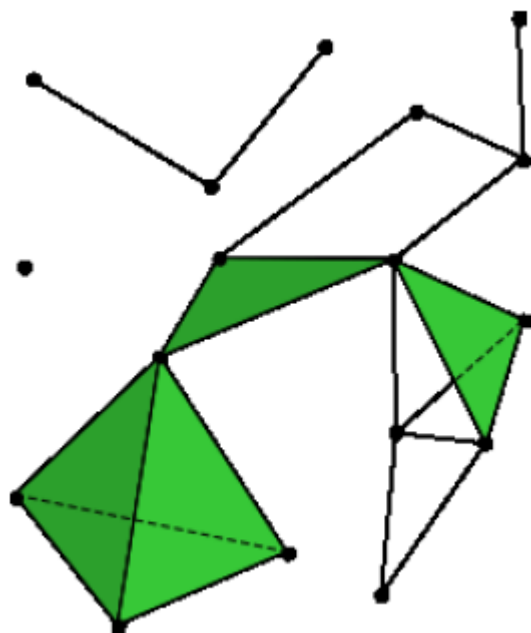


Figure 16.3. *A simplicial complex*

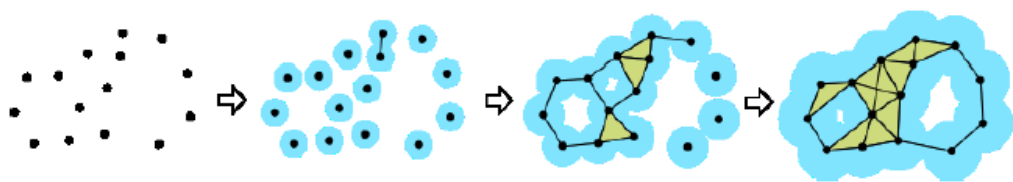


Figure 16.4. Example of a sequence of Vietoris–Rips complexes

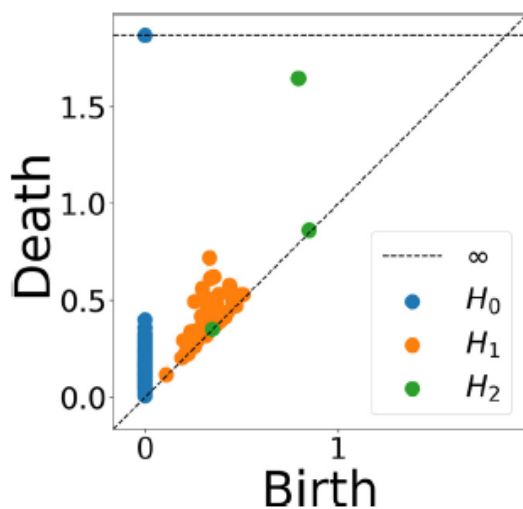


Figure 16.5. Persistence diagram of a sample of a sphere in R^3

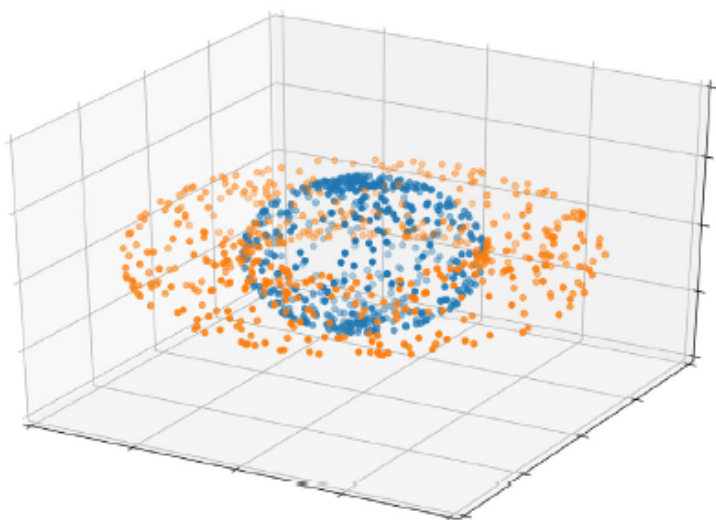


Figure 16.6. Points sampled from a torus and sphere in R^3

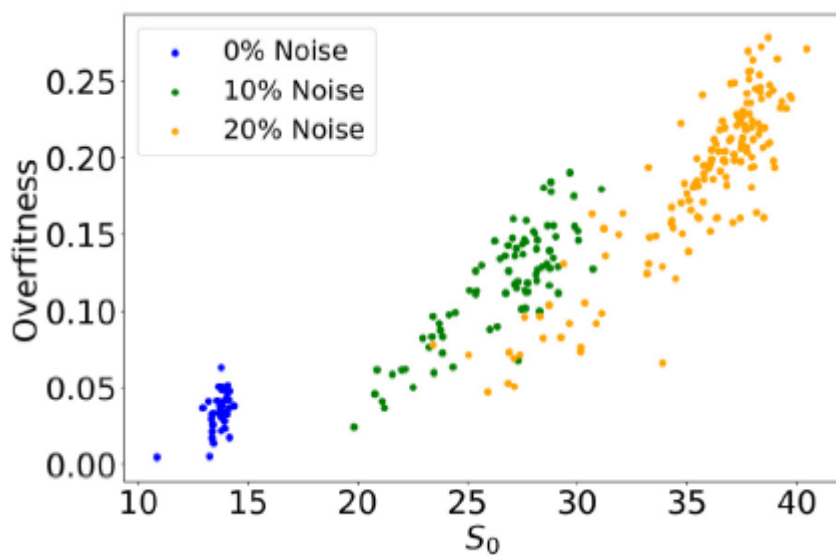


Figure 16.7. Values of S_0 from the first data set

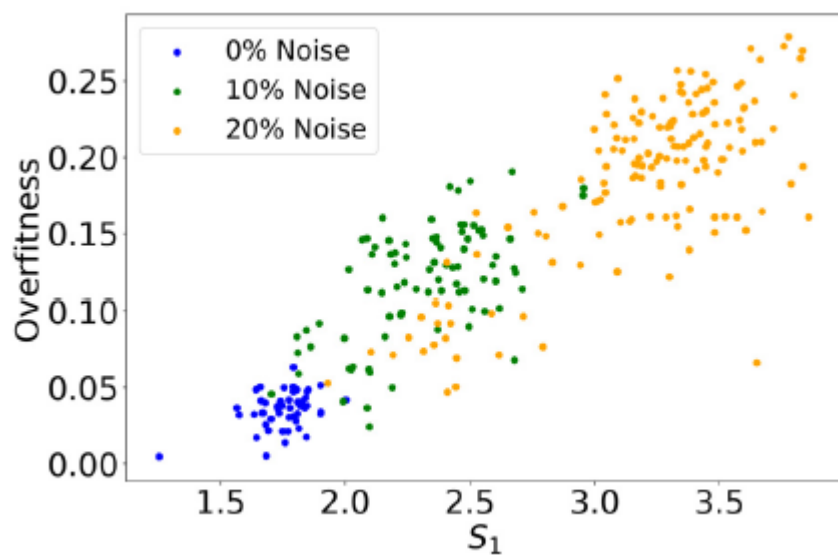


Figure 16.8. Values of S_1 from the first data set

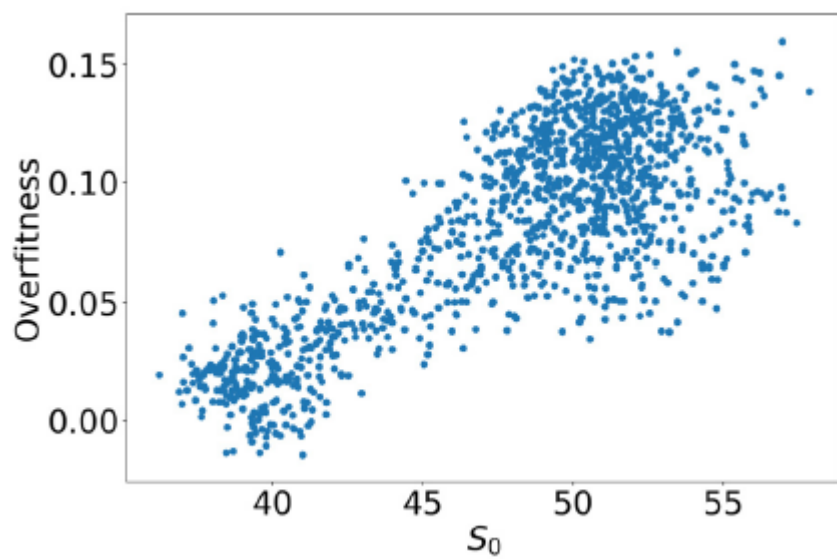


Figure 16.9. S_0 from four-dimensional data with a 30% overlap

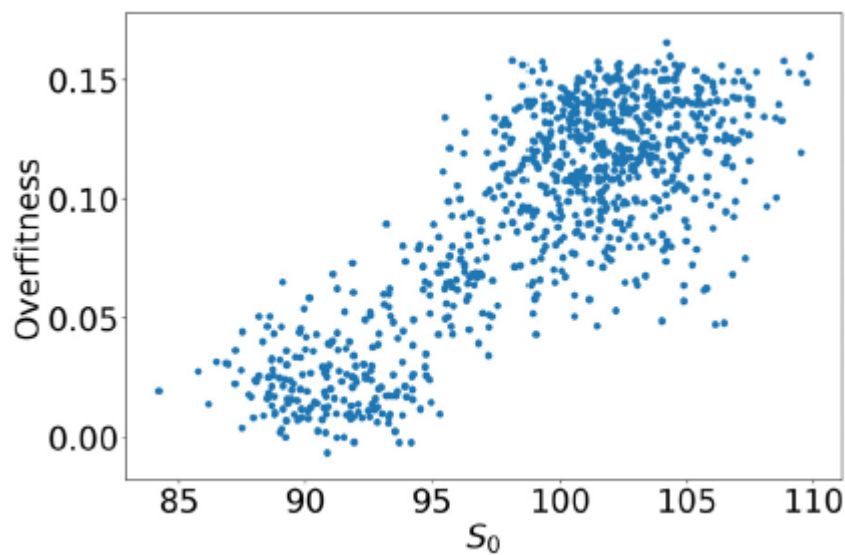


Figure 16.10. S_0 from six-dimensional data with a 30% overlap

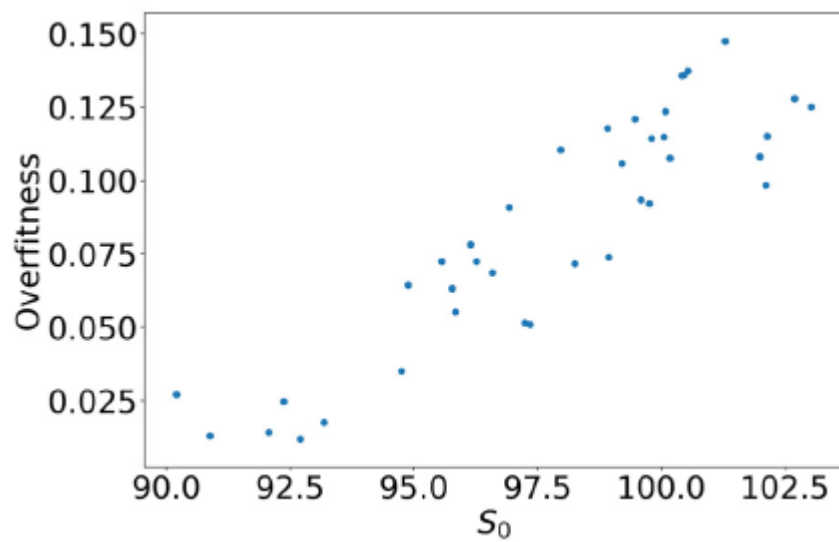
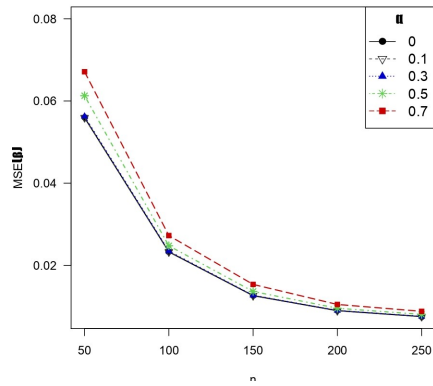
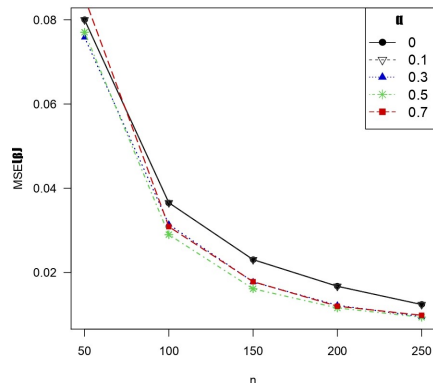


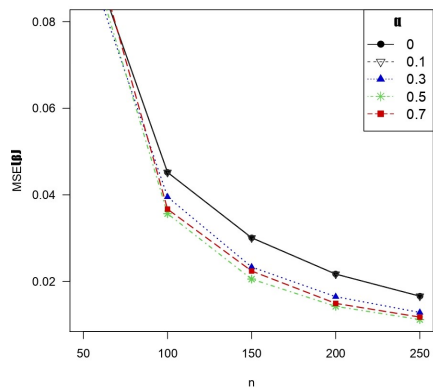
Figure 16.11. Results from an individual network with correlation $r = .904$



(a) MSE under pure data.

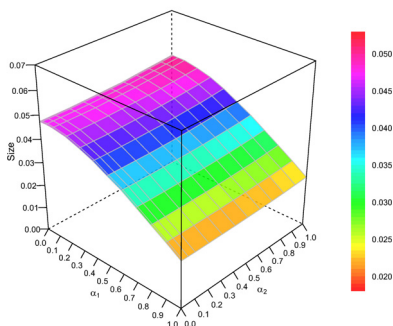


(b) MSE under 5% of contaminated data.

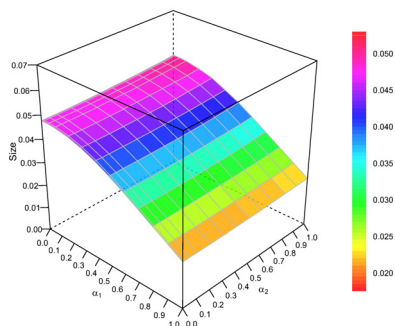


(c) MSE under 10% of contaminated data.

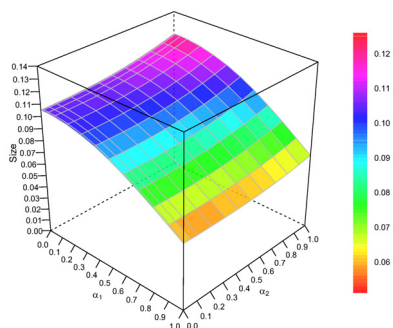
Figure 17.1. MSE in β estimation with different values of α against sample size for the Poisson regression model



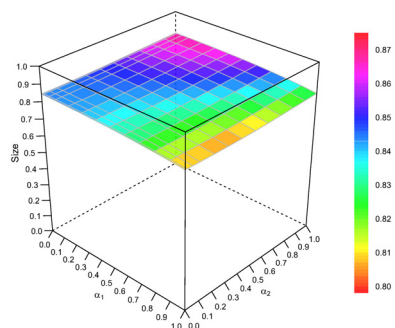
(a) $\Gamma(1)$



(b) $0.925\Gamma(1) + 0.075\Gamma(1.5)$

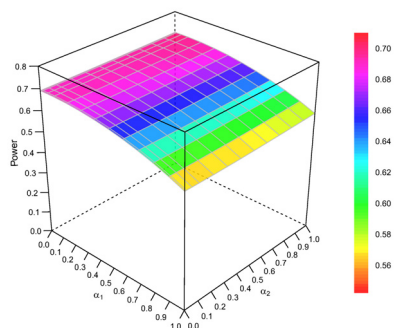


(c) $0.925\Gamma(1) + 0.075\Gamma(4)$

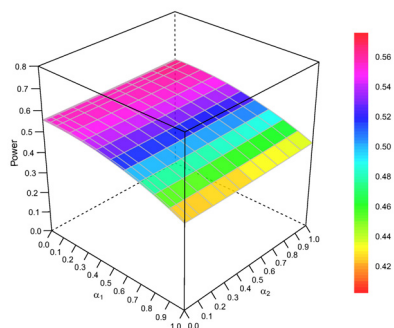


(d) $0.925\Gamma(1) + 0.075\Gamma(10)$

Figure 18.2. Size for the four contamination cases regarding the tests that can be derived from the BHHJ family

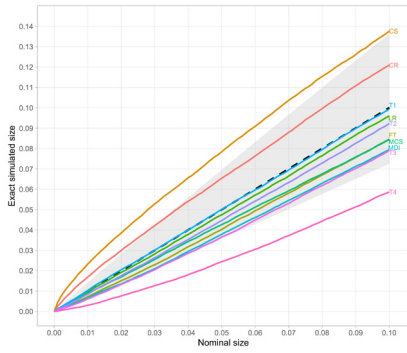


(a) $\Gamma(1.5)$

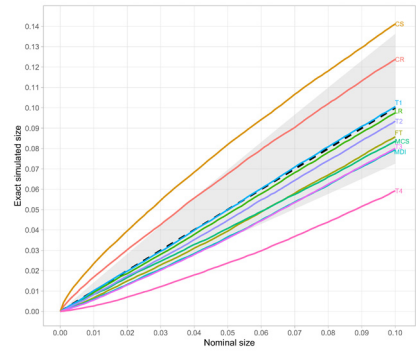


(b) $0.925\Gamma(1.5) + 0.075\Gamma(1)$

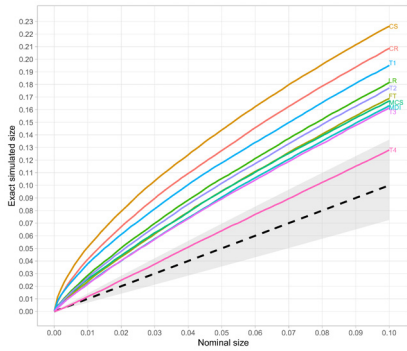
Figure 18.3. Power for the no contamination and contamination from $\Gamma(1)$ cases regarding the tests that can be derived from the BHHJ family



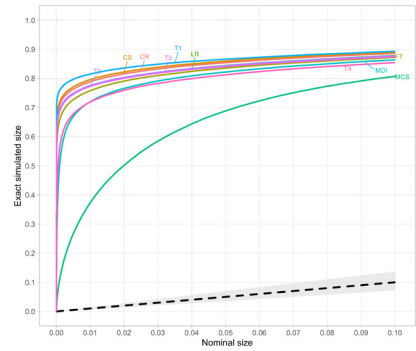
(a) $\Gamma(1)$



(b) $0.925\Gamma(1) + 0.075\Gamma(1.5)$

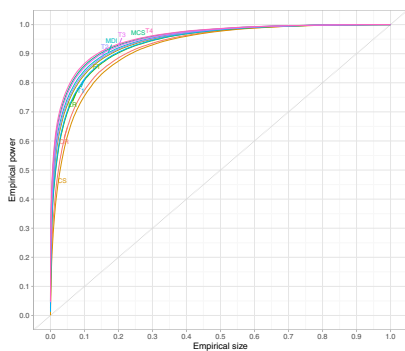


(c) $0.925\Gamma(1) + 0.075\Gamma(4)$

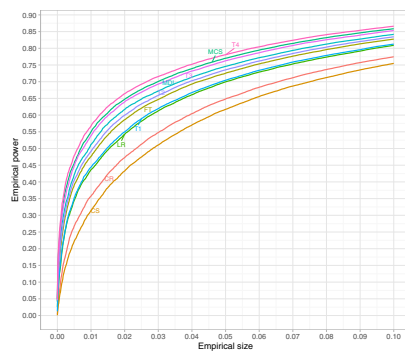


(d) $0.925\Gamma(1) + 0.075\Gamma(10)$

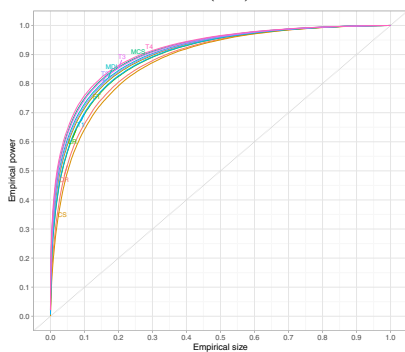
Figure 18.4. Exact simulated sizes against nominal sizes for the four cases of contamination. The gray area depicts the range of exact simulated sizes in which Dale's criterion is satisfied



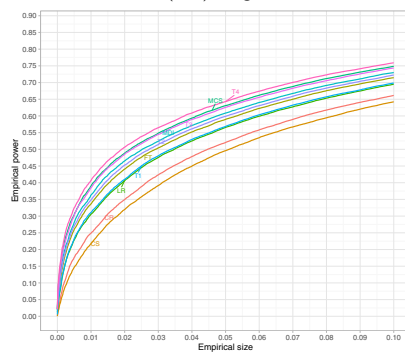
(a) $\Gamma(1.5)$



(b) $\Gamma(1.5)$ magnified



(c) $0.925\Gamma(1.5) + 0.075\Gamma(1)$



(d) $0.925\Gamma(1.5) + 0.075\Gamma(1)$ magnified

Figure 18.5. Left: empirical ROC curves for the no contamination and contamination from $\Gamma(1)$ cases. Right: the same curves magnified over a relevant range of empirical sizes

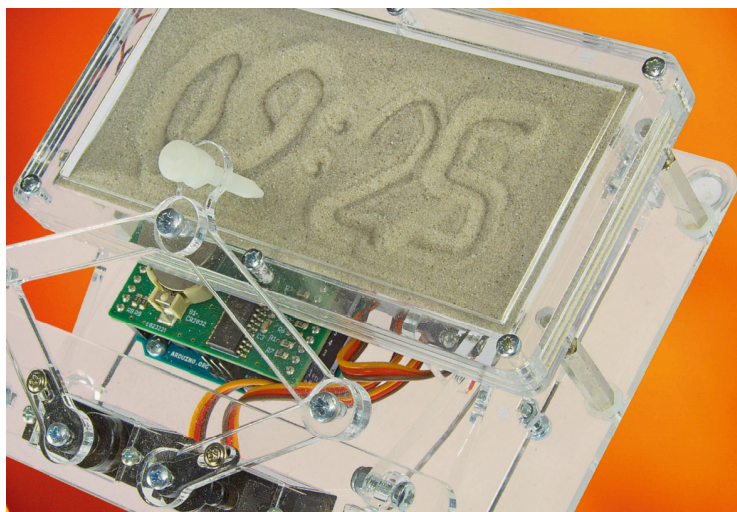


Figure 21.1. Practical implementation of the pintograph to draw figures (according to the “Sand Clock” design of (Joostens and S’heeren 2017))

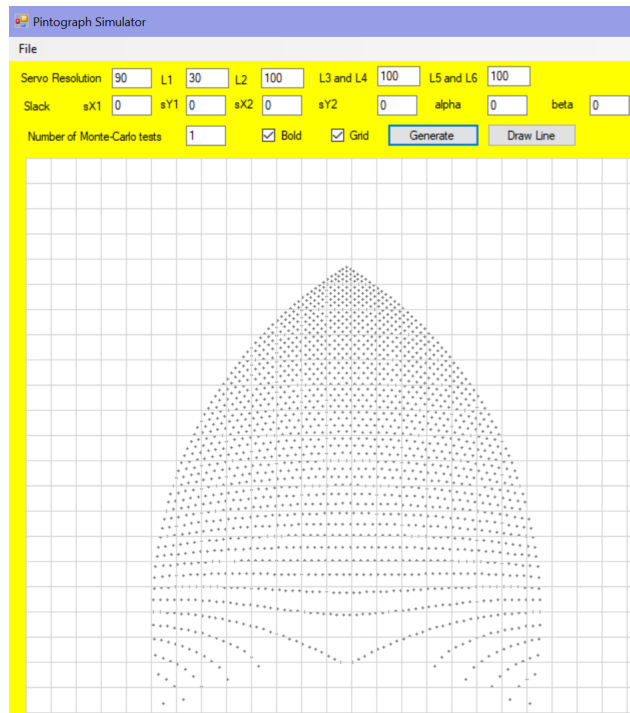


Figure 21.6. Drawing map for $L_1=30$. $L_2=L_3=L_4=L_5=L_6=100$ mm

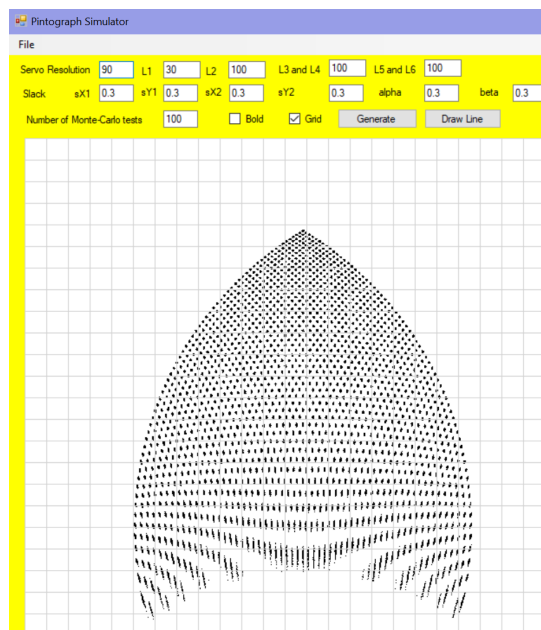


Figure 21.7. Drawing map for $L_1=30$. $L_2=L_3=L_4=L_5=L_6=100$ mm

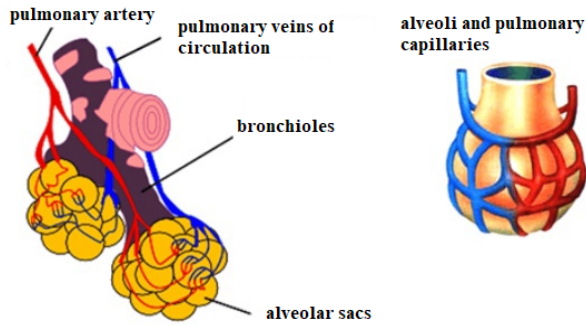


Figure 23.1. The structure of the human bronchus. Source: <https://cinetoday.ru/breast/dyhatelnaya-sistema-cheloveka-ege-chelovek-organy-sistemy-organov>

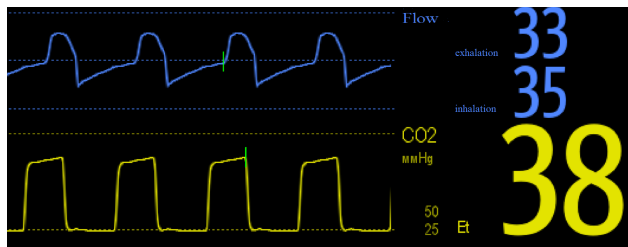


Figure 23.2. Graph of flow and capnogram

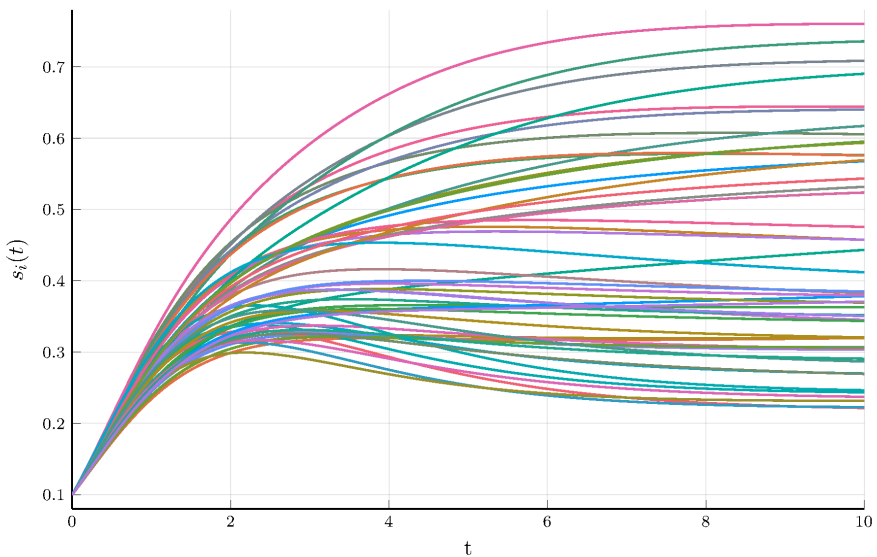


Figure 24.1. Evolution of plant sizes in a population of $N = 50$ individuals described by the Schneider system. All plants initially have the same size s_0 , and the individual traits $\theta = (x, S, \gamma)$ have a uniform distribution in a domain included within D . Trajectories are computed using the two-level numerical scheme introduced in the section

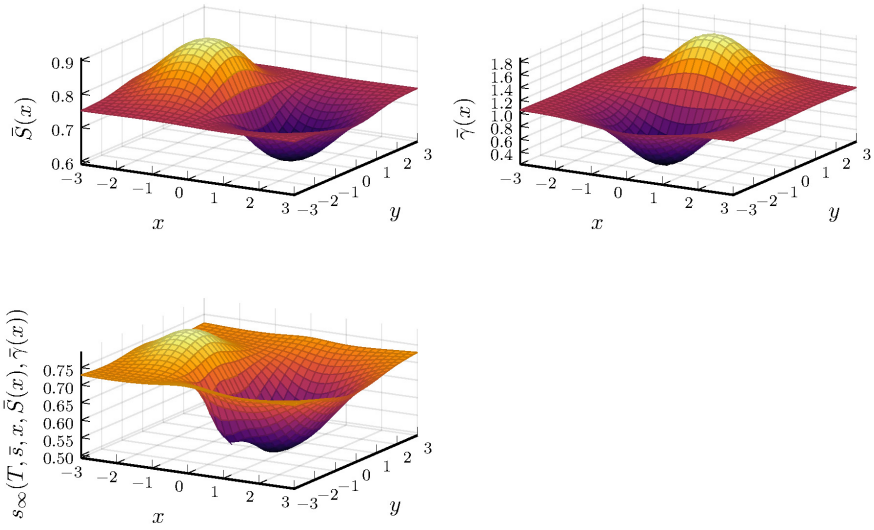


Figure 24.2. Top left: mean value of the parameter S according to the position x of the plant. Top right: mean value of the parameter γ according to the position of the plant. Bottom: evaluation of the mean-field flow s_∞ at the end of the observation period, with the mean values of parameters S and γ . s_∞ is computed using the two-level numerical scheme introduced in this section

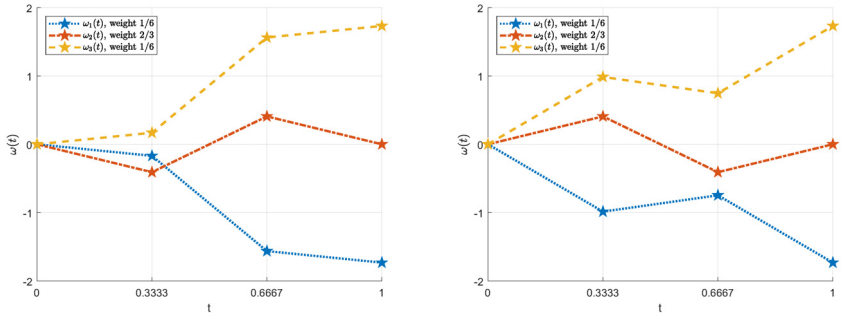


Figure 25.1. Two sets of trajectories of cubature (degree 5) on Wiener space

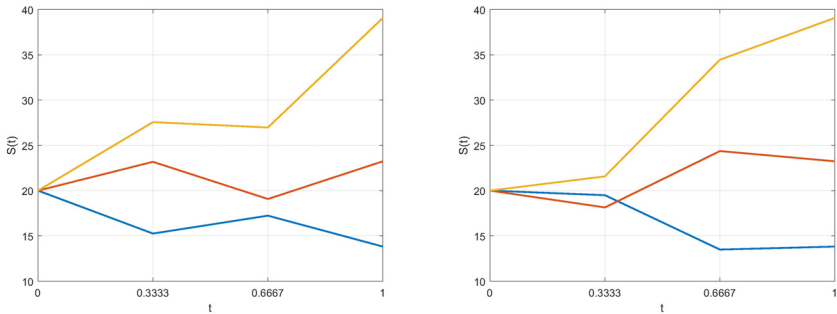


Figure 25.2. Price trajectories for $S_0 = 20$, $r = 0.12$ and $\sigma = 0.30$

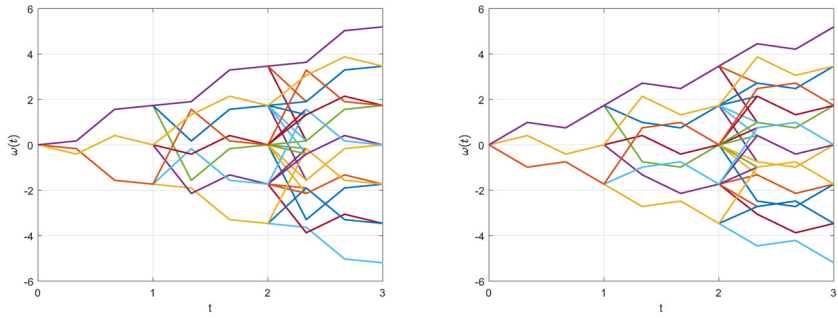


Figure 25.3. *Iterated trajectories of cubature (degree 5) on Wiener space*

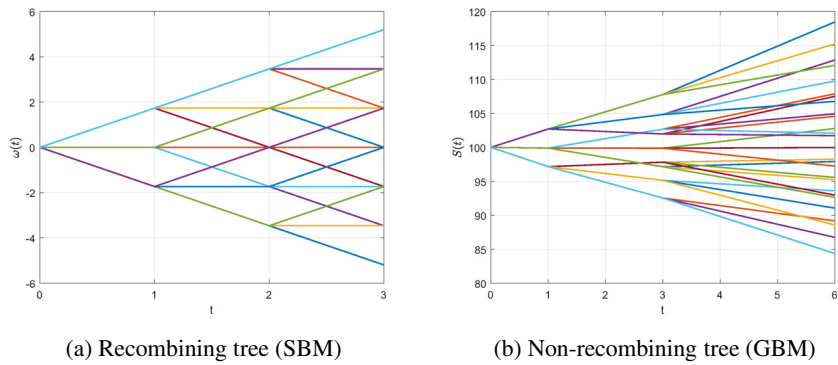


Figure 25.4. *Iterated trajectories on Wiener space*

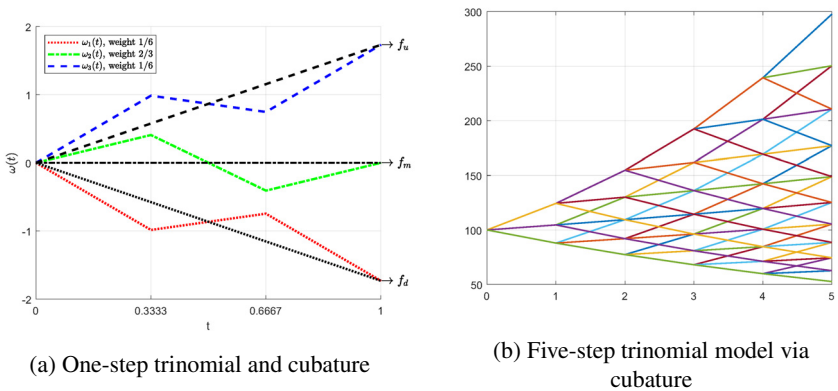


Figure 25.5. *Idea of a trinomial tree*

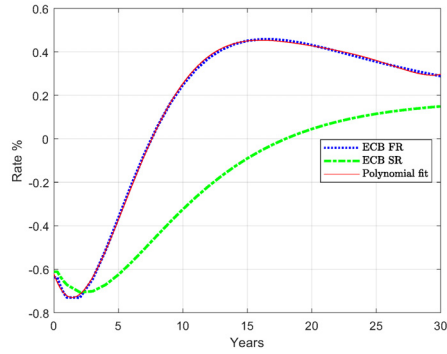
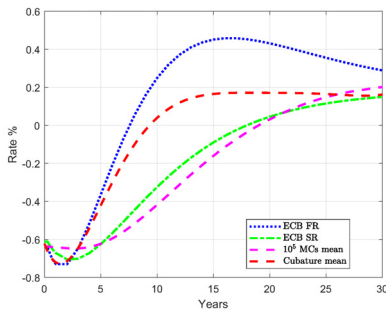
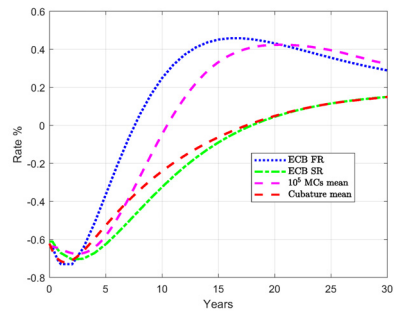


Figure 25.6. Initial FR, SR and polynomial fit

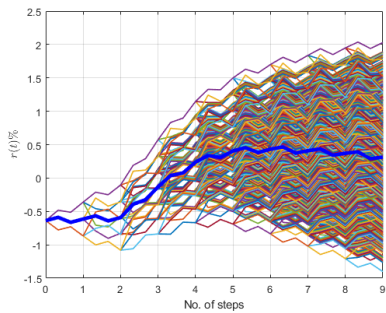


(a) $a = 0.075$ and $\sigma = 0.15$

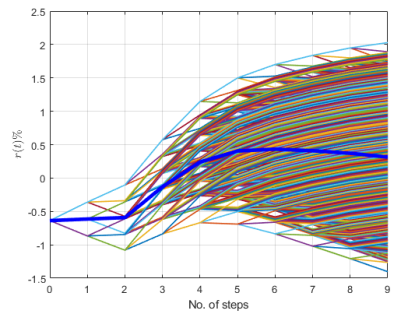


(b) $a = 0.35$ and $\sigma = 0.15$

Figure 25.7. Initial FR, SR, Monte Carlo and cubature mean

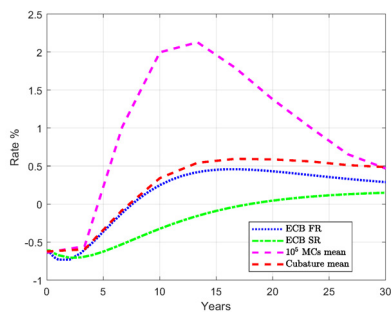


(a) Including intermediate values

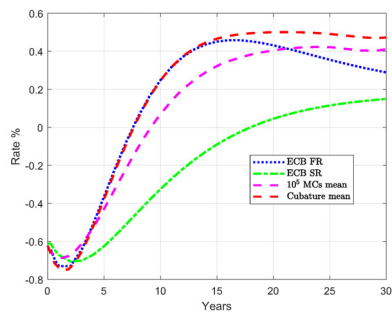


(b) Excluding intermediate values

Figure 25.8. Paths via the iterated cubature formula and their weighted average values



(a) $n = 8$, $a = 0.75$ and $\sigma = 0.15$



(b) $n = 50$, $a = 0.75$ and $\sigma = 0.15$

Figure 25.9. Effect of time discretization in MC and cubature methods

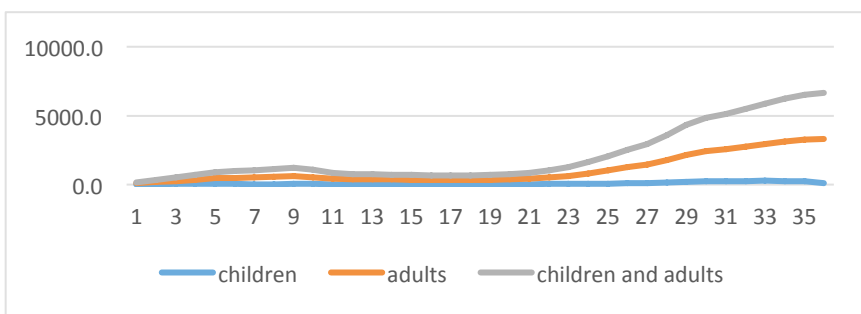


Figure 26.1. Dynamics of the number of patients with ARVI who were under observation in the district polyclinics during the period of the increase in the incidence (36 weeks of 2020, from April 20, 2020 to December 31, 2020)

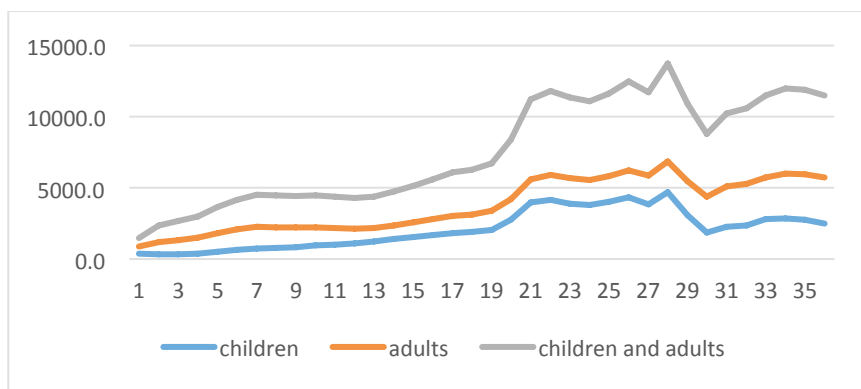


Figure 26.2. Dynamics of the number of patients with Covid-19 who were monitored in the district polyclinics during the period of the increase in the incidence (36 weeks of 2020, from April 20, 2020 to December 31, 2020)

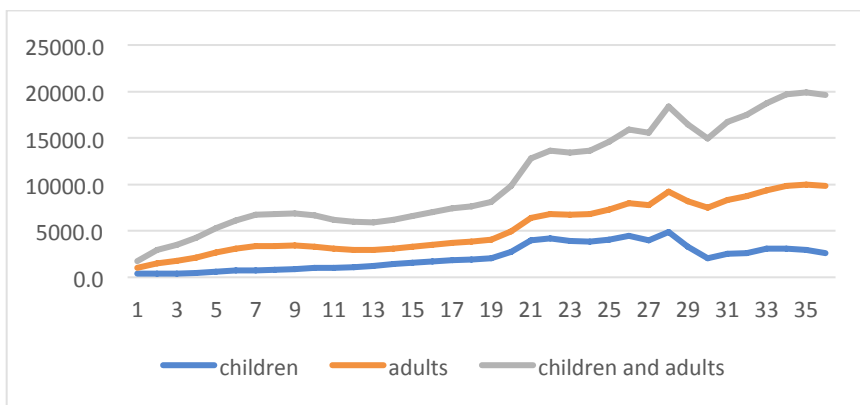


Figure 26.5. Dynamics of the number of patients with Covid-19, ARVI and community-acquired pneumonia who were monitored in the district polyclinics during the period of the increase in the incidence (36 weeks of 2020, from April 20, 2020 to December 31, 2020)