

VILNIAUS UNIVERSITETAS
MATEMATIKOS IR INFORMATIKOS FAKULTETAS

Magistro darbas

Panelinis kompiuterių skverbties modelis
Computer Penetration: a Panel Model

Sklėriūtė Indrė

VILNIUS 2007

Ekonometrinės analizės katedra

Darbo vadovas **doc. R. Lapinskas**

(parašas)

Darbas apgintas 2007 m. birželio mėn. 4 d.

Registravimo Nr. _____
2007-05-24 _____

Turinys

Santrauka	4
Summary	4
1. Įvadas	5
2. Panelinių duomenų vertinimo teorija	8
2.1. Bendras regresijos modelis	8
2.2. Panelinių duomenų analizė	11
2.2.1. Laisvasis narys	11
2.2.2. Svarių parinkimas	13
2.2.3. Atsitiktinių koeficientų modelis	14
2.2.4. Koeficientai kaip funkcijos nuo endogeninių kintamųjų	14
3. Modeliai. Tiesinis trendas	16
3.1. Modelis su tiesiniu trendu kiekvienai šaliai	16
3.2. Konstanta bendra visoms šalims	17
3.3. Fiksuotų efektų modeliai	19
3.4. Atsitiktinių efektų modeliai	20
3.5. Modelis su tiesiniu trendu ir AR nariais	23
3.6. Modelis su tiesiniu trendu ir egzogeniniais kintamaisiais	25
4. Inovacijų difuzijos modeliai	31
4.1. Kvadratinio trendo išskyrimas	32
4.2. Logistinis modelis	36
4.3. Gompertz modelis.....	40
4.3.1. Modelis su nustatomomis lubomis	40
4.3.2. Modelis, kai lubos nėra nustatomos iš anksto	47
4.4. Bass modelis.	55
5. Išvados	58
6. Literatūros sąrašas	59

Santrauka

Paskutiniais metais aktualiu klausimu tampa inovacijų modeliavimas. Pasirodant kiekvienai naujovei, verslininkams reikia suprasti ir prognozuoti rinkos vystymąsi, kad galėtų numatyti pardavimų eigą, planuoti pardavimų kiekius ir atlikti finansinį planavimą. Šiame darbe tyrinėsime kompiuterių skverbties koeficientus (penetration rate) 23 skirtingose pasaulio šalyse ir mėginsime parinkti geriausiai prognozavimui tinkantį modelį.

Skverbties koeficientas paprastai parodo prekės ženklo ar produkto kategorijos populiarumą. Jis apskaičiuojamas dalinant žmonių, kurie perka kažkokią prekę ar paslaugą, skaičių iš tos rinkos visos populiacijos skaičiaus tam tikru periodu. Taigi, skverbties koeficientas yra procentas žmonių populiacijoje, kurie per stebimą periodą nusipirko tam tikrą prekę ar paslaugą. Marketinge dažnai yra spendžiama problema: didinti pardavimus stengiantis privilioti klientus iš konkurentų ar plečiant potencialių pirkėjų populiaciją, t.y. einant į naujas rinkas.

Darbo pradžioje bus pristatyti panelinių duomenų vertinimo būdai, analizei reikalinga teorinė dalis, kai kurių reikalingų statistikų skaičiavimas. Vėliau pereisime prie pagrindinės darbo dalies - geriausio modelio parinkimo. Pirmiausia bus nagrinėjami paprasčiausi tiesinio trendo modeliai, įtraukiant autoregresijos procesą bei egzogeninius kintamuosius. Toliau įvairiais metodais mėginsime modeliuoti difuzijos procesą. Bandysime jį aproksimuoti kvadratinio trendu, taikyti logistinį bei Gompertz modelius ir galiausiai šiek tiek aptarsime Bass modelį.

Summary

Lately the modelling of innovation process is a constantly recurring problem. As with any new service for managers it is important to understand and to be able to forecast the market development for strategic, capacity and financial planning purposes. This master thesis analyses computer penetration data in 23 different countries and tries to propose the best fitting model for forecasting.

Penetration rate usually shows the popularity of a brand or a product. It is given as a percentage of the countries' households who have bought that particular brand or product at least once within a defined period of time. In marketing there is a decision to make: enlarge sales by trying to attract more costumers in the sales region or by expanding the population of potential buyers, i.e. by expanding to other emerging markets.

In the beginning of this thesis we will introduce some basic information about the panel data and some approaches of the theory needed for estimation. Later on we will pass to the main part of the work which is devoted to choosing the model. At first, we analyze some linear trend models with autoregression terms and exogenous variables. Then we examine a diffusion process by approximating it with a quadratic trend, logistic and Gompertz models. Also some basic information about the Bass diffusion model will be provided.

1. Įvadas

Kalbant apie panelinius duomenis reikėtų pradėti nuo to, kas yra sujungti (pooled) duomenys – tai bet kokių vietos ir laiko eilučių duomenų sujungimas į vieną. Jeigu kintamuosius stebime toje pačioje vietoje keliais laiko momentais, tuomet duomenys vadinami paneliniais. Panelinių duomenų analizė suteikia regresijai dvi dimensijas: vietos ir laiko.

Panelinių duomenų modeliai tampa vis populiareni taikomojoje ekonometrijos srityje dėl galimybių geriau paaiškinti žmonių elgesį. Jie naudojami daugybėje sričių. Ekonomikoje, paneliniai duomenys ypatingai dažnai naudojami tiriant įmonių veiklą bei darbuotojų atlyginimų kitimą laike. Politinių mokslų srityje galime analizuoti įvairių organizacijų bei partijų veiklą. Taip pat jie naudojami psichologijoje, sociologijoje ir sveikatos tyrimuose, analizuojant įvairias tiriamų žmonių grupių charakteristikas ir jų kitimą laike.

Dirbant su paneliniais duomenimis, mums reikia trumpesnių laiko eilučių. Tokia duomenų kombinacija gali padidinti tiek duomenų kokybę, tiek ir jų kiekį, kas būtų neįmanoma, jei naudotume tik vieno tipo duomenis, t.y. surinktus tik skirtingu laiku arba tik skirtingose vietose. Tokiu būdu galime gauti tikslesnius įverčius ir didesnės galios testų statistikas.

Panelinių duomenų aibė paprastai turi duomenų blokus, kurių kiekvienas yra sudarytas iš laiko eilučių. Mūsų duomenys susideda iš šalies pavadinimo, metų ir kompiuterių skverbties koeficiento šalyje, atitinkamais metais.

ID	TIME	PC
Azerbaijan	1997	9.00000
Azerbaijan	1998	21.00000
...
Azerbaijan	2001	24.60000
China_urban	1998	3.78000
...
UAE	2001	29.60000
USA	1993	20.97469
USA	1994	23.83626
...
USA	2004	66.02311

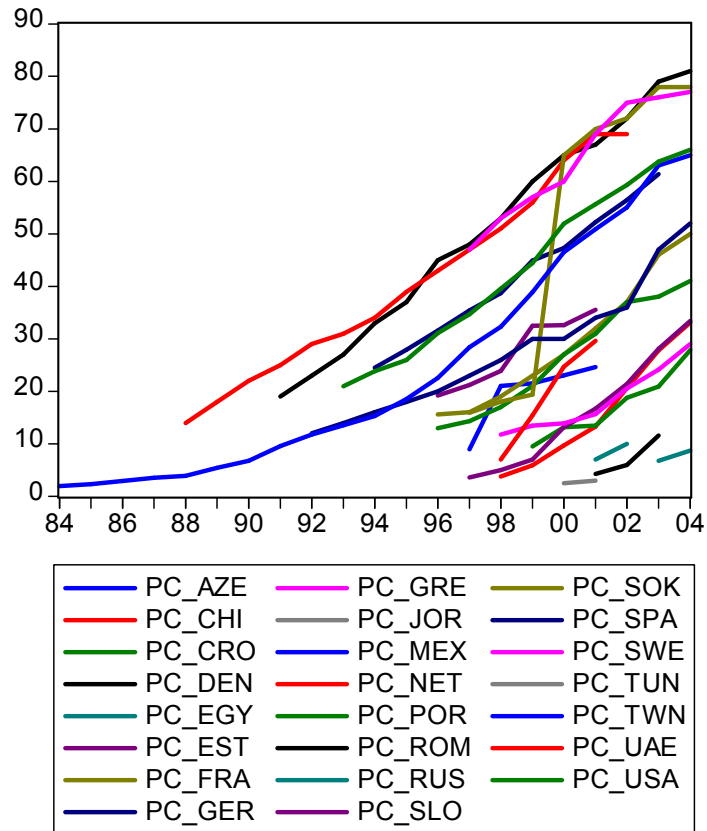
Lentelė 1: Duomenys.

Iš viso turime 23 šalių duomenis, iš kurių 12 yra Europoje, 2 Šiaurės Amerikoje, 7 Azijoje ir 2 Afrikos žemyne. Duomenys yra problematiški, kadangi šalių laiko eilučių ilgai yra labai skirtingi. Kai kurių šalių turime tik vienerių metų duomenis, pavyzdžiui, Jordanijos tik 2002, o Meksikos tik 2001. Kitų šalių, tokių kaip Taivanas, turime net 21 metų duomenis nuo 1984 iki 2004. Taigi turime nesubalansuotus panelinius duomenis, dėl to bus sunkiau atlikti kompiuterinius skaičiavimus. Naudosimės paketu E-Views, pamatysime, kad kai kurių skaičiavimų jis iš tiesų atlikti negalės (tokių kaip SUR (Seemingly Unrelated Regression) metodo taikymas nesubalansuotiems paneliniams duomenims).

EUROPA	Š. AMERIKA	AZIJA	AFRIKA
Danija	JAV	Azerbaidžianas	Egiptas
Estija	Meksika	JAE	Tunisas
Graikija		Jordanija	
Ispanija		Kinija	
Kroatija		Rusija	
Olandija		Pietų Korėja	
Portugalija		Taivanas	
Prancūzija			
Rumunija			
Slovėnija			
Švedija			
Vokietija			

Lentelė 2: Šalys, kurių kompiuterių skverbties duomenis turime.

Pabraižykime grafiką, kad matytume bendras tendencijas. Kaip matome grafike, easnčiame apačioje, kompiuterių pardavimai prasideda skirtingais laiko momentais. Iš tiesų dažniausiai skirtingose šalyse inovacijos atsiranda skirtingu metu, tačiau skverbties dinamika būna panaši. Darydami skirtingas prielaidas, galime gauti skirtingus modelius. Aptarti tinkamus modelius yra šio darbo tikslas. Tolesniuose skyriuose bus išdėstyta modelių vertinimo teorija, jų aprašymas ir skaičiavimų rezultatai bei prognozės. Modelio koeficientus vertinsime remdamiesi visu stebimu periodu, t.y. nuo 1984 iki 2004, o prognozes skaičiuosime remdamiesi duomenimis nuo 2000 metų, nes kai kurioms analizuojamoms šalims, tai yra pirmasis duomuo, be to, manoma, kad norint paskaičiuoti teisingas prognozes, reikėtų remtis maždaug trečdaliu visos imties, mūsų atveju tai būtų maždaug 1998 metai. Prognozuosime 10 metų į priekį, ekonomikoje tai yra laikoma ilgalaikė prognoze. Šios prognozės reikalingos strateginiam planavimui, o tai nebūna trumpalaikiai uždaviniai.



Grafikas 1: Šalys, kurių kompiuterių skverbties duomenis turime.

2. Panelinių duomenų vertinimo teorija

Šiame skyriuje aprašysime bendrą regresijos modelių vertinimo teoriją, pagrindines statistikas, bei pagrindinius panelinių duomenų vertinimo metodus. Panelinių duomenų vertinimas pirmiausia susijęs su laisvojo nario parinkimu, jis gali būti bendras visoms šalims (common), fiksuotų efektų, kai kiekviena šalis turi savo laisvąjį narį ir atsitiktinių efektų, kai tariama, kad laisvieji nariai gali būti traktuojami kaip nepriklausomi atsitiktiniai kintamieji su nuliniu vidurkiu ir baigtine dispersija kiekvienai šaliai. Kitas aspektas yra parinkti tinkamus svorius atsižvelgiant į modelio liekanų heteroskedastiškumą bei koreliaciją.

2.1. Bendras regresijos modelis

$$y_{it} = \alpha + x_{it}'\beta_i + \delta_i + \gamma_t + \varepsilon_{it}, \quad i = \text{country}, \quad i = 1, 2, \dots, N, \quad t = \text{time}, \quad t = 1, 2, \dots, T$$

Čia y_{it} yra endogeninis kintamasis, x_{it} - k-matis regresorių vektorius, ε_{it} - modelio liekanos. α yra bendrasis modelio laisvasis narys, δ_i - specifinis koeficientas tarp šalių (cross-section), o γ_t - koeficientas, kintantis tarp laikotarpių (period specific). Mūsų atveju aktualu bus įvertinti δ_i , bei, žinoma, β_i .

Paprasčiausias panelinių duomenų vertinimo metodas yra mažiausių kvadratų. Jo principas – parinkti tokią tiesę, kad gautų nuokrypių kvadratų suma būtų mažiausia.

Regresijos koeficientų skaičiavimas MKM

Tarkim turime paprastą regresinį modelį, užrašytą matricine forma:

$$Y = X\beta + \varepsilon$$

čia Y yra T -dimensinis vektorius, kuris turi endogeninio kintamojo reikšmes. X – $T \times k$ egzogeninių kintamųjų matrica. β - k -dimensinis koeficientų vektorius, ε - T -dimensinis paklaidų vektorius. T – stebėjimų skaičius, o k – kintamųjų skaičius.

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_T \end{pmatrix}; \beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}; \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_T \end{pmatrix}; X = \begin{pmatrix} 1 & X_{21} & \dots & X_{k1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{2T} & \dots & X_{kT} \end{pmatrix} = \begin{pmatrix} X_1^T \\ \vdots \\ X_k^T \end{pmatrix}$$

Taigi vektorinio parametro β mažiausių kvadratų įvertis yra toks:

$$\hat{\beta} = \arg \min_{b \in R^k} f(b)$$

Empiriškai koeficientai apskaičiuojami pagal tokią MKM formulę:

$$\hat{\beta} = (X^T X)^{-1} X^T Y, \text{ čia } X^T - \text{transponuota matrica.}$$

Paprasto tiesinio regresinio modelio atveju, koeficientai išmatuoja marginalinę egzogeninio kintamojo įtaką endogeniniam kintamajam, kai visi likę kintamieji yra fiksuoti. Įvertinę koeficientus suskaičiuojame modelio liekanas - modelio nuokrypius nuo endogeninio kintamojo.

Vertinant koeficientus paskaičiuojamos jų statistinės charakteristikos. Apie jas šiek tiek plačiau.

Standartinės paklaidos

Standartinės paklaidos – tai standartiniai nuokrypiai. Šis dydis atspindi įvertintų koeficientų patikimumą. Kuo didesnis šis dydis, tuo daugiau atsitiktinumo, triukšmo juose yra.

Paklaidos dispersija $D\varepsilon = \sigma^2$ vertinama: $\widehat{\sigma}^2 = \hat{\varepsilon}^T \hat{\varepsilon} / (T - k)$; $\hat{\varepsilon} = Y - X\hat{\beta}$

Koeficientų įverčių vektoriaus $\hat{\beta}$ kovariacijos matrica vertinama reiškiniu:

$$\widehat{\text{cov}}(\hat{\beta}) = \widehat{\sigma}^2 (X^T X)^{-1},$$

o šios matricos narių ant įstrižainės kvadratinės šaknys yra koeficientų įverčių standartinės paklaidos:

$$se(\hat{\beta}_i) = \sqrt{\text{var}(\hat{\beta}_i)}, \quad i = 1, \dots, k.$$

t-reikšmė

t-reikšmė pateikiama visiems parametrams ir parodo kintamojo reikšmingumą. Ši statistika patikrina hipotezę, kad $H_0: \beta_i = 0$ su alternatyva, jog $H_1: \beta_i \neq 0$, $i = 1, \dots, k$. Reikšmės gaunamos taip:

$$t = \frac{\hat{\beta}_i}{se(\hat{\beta}_i)}, \quad i = 1, \dots, k.$$

p-reikšmė

p-reikšmės stulpelyje pateiktos tikimybės, kad su pasirinktu reikšmingumo lygmeniu koeficientas prie to kintamojo gali būti lygus nuliui. Šios reikšmės atitinka hipotezę, kad koeficientas lygus nuliui, kai p-reikšmė yra didesnė už pasirinktą reikšmingumo lygmenį, pvz.: 0.05.

Taip pat vertinant modelį žiūrime į įvertintas modelio statistikas. Šiek tiek apie jas.

R² - determinacijos koeficientas

R² (determinacijos koeficientas) parodo, kaip sėkmingai regresijos lygtimi prognozuojamos endogeninio kintamojo reikšmės, remiantis turimais duomenimis. Jei ši reikšmė artėja prie 1, tai modelis endogeninį kintamąjį paaiškina idealiai ir regresijos lygtis tinka kuo puikiau. Taigi kuo ji arčiau vieneto, tuo modelis geresnis.

R² (determinacijos koeficientas) apskaičiuojamas taip:

$$R^2 = 1 - \frac{\hat{\varepsilon}^T \hat{\varepsilon}}{(Y - \bar{Y})^T (Y - \bar{Y})}; \quad \bar{Y} = \sum_{t=1}^T Y_t / T, \quad \text{čia } \bar{Y} - \text{endogeninio kintamojo vidurkis.}$$

R^2 skaičiavimas paneliniuose modeliuose pagrįstas skirtumu tarp kvadratinių liekanų sumos (Residual Sum of Squares - RSS) iš įvertinto panelinio (fiksuoatų ar atsitiktinių efektų) modelio ir kvadratinių liekanų sumos (RSS) iš modelio, kuris vertinamas įtraukus tik bendrą (common) konstantą, o ne fiksuotus ar atsitiktinius efektus. Tuomet šios statistikos interpretacija yra tokia, kad ji paaiškina modelio specifikacijos galią, įtraukiant ir fiksuotus ar atsitiktinius efektus. Ši statistika kaip parametrų skaičių naudoja įvertintų koeficientų skaičių, įtraukiant ir efektus.

R^2 (adjusted)

R^2 (adjusted) skiriasi nuo R^2 tik tuo, kad pastarasis dydis nemažėja, kai į modelį įtraukiami nereikšmingi kintamieji, R^2 (adjusted) reaguoja į šiuos nesvarbius kintamuosius, neturinčius įtakos paaiškinamajam dydžiui. R^2 (adjusted) kartais dar žymimas \bar{R}^2 .

Ši statistika skaičiuojama taip:

$$\bar{R}^2 = 1 - (1 - R^2) \frac{T-1}{T-k}$$

Galime pastebėti, kad R^2 (adjusted) niekada nebūna didesnis už R^2 , ir tuo atveju, kai egzogeniniai kintamieji silpnai paaiškina endogeninį, jis gali įgyti net neigiamą reikšmę.

Kvadratinių liekanų suma (Residual Sum of Squares - RSS)

Kvadratinių liekanų suma gaunama taip: $RSS = \sum_{i=1}^T \varepsilon_i^2$. Ši reiškinį padalinę iš $(T-k)$, gautume

paklaidų dispersijos įvertį σ^2 . Be to, ištraukę iš šio skaičiaus šaknį, gautume regresijos standartinės paklaidos įvertį (angl. S.E. of Regression).

F-statistika

F-statistika tikrinama nulinė hipotezė $H_0 : \beta_2 = \beta_3 = \dots = \beta_k = 0$, kai alternatyva H_1 : bent vienas iš šių koeficientų nėra lygus 0. Kai nulinė hipotezė teisinga, F-statistika turi Fišerio skirstinį su $(k-1, T-k)$ laisvės laipsniais. Kai F-statistikos reikšmės gana didelės, nulinę hipotezę atmetame ir atvirkščiai. Taip pat prie šios statistikos dažnai pateikiamos tikimybės. Kai tikimybė mažesnė už 0,05 (ar kitą pasirinktą reikšmingumo lygmenį), nulinę hipotezę atmetame.

Durbin – Watson

Ši statistika atlieka testą, tikrinantį pirmos eilės serijinę koreliaciją. Kitaip tariant, Durbin – Watson tikrina, ar yra tiesinė priklausomybė tarp greta esančių regresijos įvertintų liekanų. Tikrinama hipotezė, kad $\rho = 0$, kai

$$u_t = \rho u_{t-1} + \varepsilon_t.$$

Jei nėra serijinės koreliacijos, tuomet statistikos reikšmė yra apie 2. Jei artėja prie nulio – turime teigiamą serijinę koreliaciją, jei prie 4 – neigiamą serijinę koreliaciją.

Reikėtų paminėti, kad, jei dešinėje regresijos pusėje yra endogeninių kintamųjų su pavėlinimu, tuomet testas laikomas neveiksmingu. Dar vienas testo trūkumas yra tas, kad hipotezė yra tik viena: pirmos eilės serijinė koreliacija yra arba jos nėra.

Paneliniuose modeliuose ši statistika skaičiuojama paėmus visų pirmos eilės liekanų koreliaciją.

RMSE ir MAE

Jos bus naudojamos geresniam modelių parinkimui, skaičiuojamos bus kiekvienai šaliai. Šaknis iš kvadratinių vidutinių paklaidų (Root Mean Square Error - RMSE) ir vidutinės absoliučios paklaidos (Mean Absolute Error – MAE):

$$RMSE = \sqrt{\frac{\varepsilon_1^2 + \dots + \varepsilon_N^2}{N}} \quad \text{ir} \quad MAE = \frac{|\varepsilon_1| + \dots + |\varepsilon_N|}{N}.$$

2.2. Panelinių duomenų analizė

Paketas E-Views mums leidžia panelinį modelį įvertinti keliais būdais, keičiant laisvojo nario specifikaciją bei parenkant svorius, atsižvelgiant į liekanų heteroskedastiškumą ar koreliaciją. Taigi, mūsų vertinamas modelis:

$$y_{it} = \alpha + x_{it}'\beta_i + \delta_i + \varepsilon_{it}, \quad i = \text{country}, \quad i = 1, 2, \dots, N, \quad t = \text{time}, \quad t = 1, 2, \dots, T$$

Toliau pakalbėkime apie jo vertinimo būdus plačiau.

2.2.1. Laisvasis narys

1. Jo gali nebūti: $\alpha = 0$.

2. Jis gali būti bendras visoms šalims: α .

3. Modelis gali būti fiksuotų efektų - laisvasis narys gali būti skirtingas kiekvienai šaliai, t.y. kiekvienai šaliai atskirai įvertinamas δ_i . Pirmiausia suskaičiuojamas kintamųjų vidurkis (tiek endogeninio, tiek egzogeninių) ir atimamas iš kiekvienos tos šalies kintamojo reikšmės, tuomet koeficientai vertinami MKM, naudojant transformuotus duomenis:

$$y_i - \bar{y}_i = (x_i - \bar{x}_i)' \beta + (\varepsilon_i - \bar{\varepsilon}_i), \quad \text{kur} \quad \bar{y}_i = \sum_t \frac{y_{it}}{N}, \quad \bar{x}_i = \sum_t \frac{x_{it}}{N}, \quad \bar{\varepsilon}_i = \sum_t \frac{\varepsilon_{it}}{N}.$$

Kovariacinės matricos įverčiai gaunami skaičiuojant pagal įprastą mažiausių kvadratų kovariacijos formulę, ją pritaikant duomenims, iš kurių atimtas vidurkis:

$\text{var}(b_{FE}) = \hat{\sigma}_W^2 (\tilde{X}' \tilde{X})^{-1}$, kur \tilde{X} atspindi X atėmus vidurkį (FE žymi fiksuotus efektus) ir

$$\hat{\sigma}_W^2 = \frac{e_{FE}' e_{FE}}{NT - N - T} = \frac{\sum_{it} (\tilde{y}_{it} - \tilde{x}_{it}' b_{FE})^2}{NT - N - T},$$

čia $e_{FE}' e_{FE}$ yra kvadratinių liekanų suma (RSS) iš fiksuotų efektų modelio. Jei duomenys yra nesubalansuoti, tuomet NT yra keičiama į visų turimų reikšmių skaičių. Fiksuoti efektai – laisvieji nariai kiekvienai šaliai - nėra vertinami tiesiogiai, jie įvertinami pagal formulę:

$$\hat{\alpha}_i = \sum_i (\bar{y}_i - \bar{x}_i' b_{FE}) / N.$$

Reikia pastebėti, kad rezultatų lentelėje pateikiamas laisvasis narys, o fiksuoti efektai pateikiami taip, kad jų suma būtų lygi maždaug nuliui. Taigi jie interpretuojami kaip kiekvienos šalies nukrypimai nuo bendro vidurkio. Ar fiksuoti efektai nėra pertekliniai tikrinsime naudodami kvadratų sumas (sums-of-squares (F-test)) ir tikėtinumo funkciją (likelihood function (Chi-square test)).

4. Modelis gali būti atsitiktinių efektų – laisvieji nariai gali būti traktuojami kaip nepriklausomi atsitiktiniai kintamieji su nuliniu vidurkiu ir baigtine dispersija kiekvienai šaliai. Svarbi atsitiktinių koeficientų specifikacija yra ta, kad atsitiktiniai efektai δ_i ir modelio liekanos ε_{it} yra nekoreliuoti, t.y. $E(\delta_i \varepsilon_{it}) = 0$.

Atsitiktinių efektų modeliai vertinami apibendrintų mažiausių kvadratų metodu (generalized least squares - GLS) tokiais žingsniais:

1. naudojame liekanas e_{FE} iš fiksuotų efektų modelio, kad įvertinti ε_{it} dispersiją ir kovariaciją, naudojant $\hat{\sigma}_W^2$, kuris aprašytas aukščiau.
2. Įvertiname atskirai modelius kiekvienai šaliai, paskaičiuojame vidurkį ir apskaičiuojame:

$$\hat{\sigma}_B^2 = \frac{e_B' e_B}{N - K}, \quad \hat{\sigma}_u^2 = \hat{\sigma}_B^2 - \frac{\hat{\sigma}_W^2}{T},$$

čia $e_B' e_B$ yra kvadratinės liekanos iš tarpgruvinės (B – between-group) regresijos.

3. Taikome mažiausių kvadratų metodą GLS transformuotiems kintamiesiems (X susideda iš konstantos ir regresorių x):

$$y_{it}^* = y_{it} - \hat{\lambda} \bar{y}_i, \quad X_{it}^* = X_{it} - \hat{\lambda} \bar{X}_i, \quad \text{kur } \hat{\lambda} = 1 - \hat{\sigma}_W / \hat{\sigma}_B.$$

Atsitiktinių efektų įverčiai paskaičiuojami: $\hat{u}_i = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_B^2} \left(\sum_i (y_{it} - X_{it} b_{RE}) \right)$.

4. Modelio statistikos skaičiuojamas duomenims be svorių ir su svoriais. Su svoriais paskaičiuojama iš GLS, be svorių: $\varepsilon_{it} = y_{it} - X_{it} b_{RE} - \hat{u}_i$.

Pagrindinė atsitiktinių efektų modelio vertinimo prielaida yra, kad atsitiktiniai efektai yra nekoreliuoti su paaiškinančiaisiais kintamaisiais. Pagrindinis šios prielaidos tikrinimas yra **Hausman testas**, kuris lygina fiksuotus ir atsitiktinius efektus. Norėdami parinkti geresnį modelį šiame darbe jį ir naudosime. Statistika skaičiuojama taip:

$$m = (b_{FE} - b_{RE})' \hat{\Sigma}^{-1} (b_{FE} - b_{RE}) \sim X^2(k), \quad \text{kur } \hat{\Sigma} = \text{Var}(b_{FE} - b_{RE}) = \text{Var}(b_{FE}) - \text{Var}(b_{RE}),$$

čia FE žymi fiksuotus efektus, o RE – atsitiktinius efektus.

2.2.2. Svių parinkimas

Modelis vertinamas sujungtų mažiausių kvadratų metodu (Pooled Least Squares - PLS). Gali būtų daromos skirtingos prielaidos apie modelio paklaidų koreliuotumą bei dispersiją. Panagrinėsime skirtingus vertinimo variantus. Svoriai bus taikomi modelius vertinant fiksuotų efektų metodu, nes atsitiktinių efektų modelyje E-Views svorių naudoti neleidžia.

1. Kiekvienam stebėjimui duodamas vienodas svoris (no weighting). Tai paprasčiausias metodas, jis taikomas, kai liekanos yra homoskedastinės ir nekoreliuotos. Kovariacinė liekanų matrica:

$$\Omega = E\varepsilon\varepsilon' = E \begin{pmatrix} \varepsilon_1\varepsilon_1' & \varepsilon_2\varepsilon_1' & \dots & \varepsilon_N\varepsilon_1' \\ \varepsilon_1\varepsilon_2' & \varepsilon_2\varepsilon_2' & \dots & \varepsilon_N\varepsilon_2' \\ \dots & \dots & \dots & \dots \\ \varepsilon_1\varepsilon_N' & \varepsilon_2\varepsilon_N' & \dots & \varepsilon_N\varepsilon_N' \end{pmatrix}, \quad \Omega = \sigma^2 I_N \otimes I_T$$

2. Svoriai tarp šalių (cross-section weights). Tariama, kad kovariacija tarp skirtingų šalių liekanų yra lygi nuliui. Taigi tariama, kad $E(\varepsilon_{it}\varepsilon_{it} | X_i^*) = \sigma_i^2$ ir $E(\varepsilon_{is}\varepsilon_{jt} | X_i^*) = 0$, visiems i, j, s, t , kur $i \neq j$ ir $s \neq t$, be to, X_i^* sudarytas iš X_i ir δ_i , jei modelis įvertintas su fiksuotais efektais. Naudojamos įvertintų liekanų dispersijos ir pagal jas parenkami svoriai. Svoriai reikalingi tuomet, jei regresijos funkcijos liekanos yra heteroskedastinės tarp šalių ir nekoreliuotos. Naudojami šie įvertinimo žingsniai:

1. Kiekviena lygtis įvertinama atskirai, naudojant mažiausių kvadratų metodą.
2. Išsaugomos paklaidų reikšmės iš 1 žingsnyje gautų įvertinimų.
3. Paklaidos naudojamos įvertinti paklaidų dispersijas regresijos lygtims.
4. Įvertintos paklaidų dispersijos naudojamos Apibendrintų Mažiausių Kvadratų (Generalized Least Squares - GLS) procedūroje, įvertinant regresijos lygtis.

Taigi liekanų kovariacinė matrica yra tokio pavidalo:

$$\Omega = E\varepsilon\varepsilon' = E \begin{pmatrix} \sigma_1^2 I_{T_1} & 0 & \dots & 0 \\ 0 & \sigma_2^2 I_{T_2} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_N^2 I_{T_N} \end{pmatrix};$$

čia dispersija paskaičiuojama $\hat{\sigma}_i^2 = \sum_{t=1}^{T_i} (y_{it} - \hat{y}_{it})^2 / T_i$, kur Y yra įvertinti mažiausių kvadratų metodu.

3. SUR (Seemingly Unrelated Regression) – įvertinimui naudojama įvertintų liekanų kovariacijos matrica.

SUR yra aibė regresinių lygčių, kurios atrodo nesusijusios, tačiau iš tikrųjų yra susijusios. SUR leidžia paklaidų koreliuotumą bei heteroskedastiškumą tarp atskirų, tačiau tarpusavyje susijusių regresijos lygčių. Taigi, leisdami paklaidų koreliuotumą bei heteroskedastiškumą, galime pagerinti koeficientų įverčius. Gali būti, kad konkrečiais metais vyko kažkokie pokyčiai, kurie turėjo įtakos visoms šalims. Tas pokytis nebuvo įvertintas vertinant atskiras lygtis, dėl to atsispindėjo paklaidose. SUR procedūra naudoja šią informaciją koeficientų įverčių pagerinimui. Taigi tariama,

kad $E(\varepsilon_{it}\varepsilon_{jt} | X_t^*) = \sigma_{ij}$ ir $E(\varepsilon_{is}\varepsilon_{jt} | X_t^*) = 0$, visiems i, j, s, t , kur $s \neq t$. Svarbu, kad kovariacija neleidžiama tarp laiko periodų, o tik tarp šalių. Naudojami aukščiau išvardinti įvertinimo žingsniai, tik šį kartą vertinamos ir kovariacijos.

Liekanų kovariacinė matrica yra:

$$\Omega = E\varepsilon\varepsilon' = E \begin{pmatrix} \sigma_{11}I_T & \sigma_{12}I_T & \dots & \sigma_{1N}I_T \\ \sigma_{21}I_T & \sigma_{22}I_T & \dots & \sigma_{2N}I_T \\ \dots & \dots & \dots & \dots \\ \sigma_{N1}I_T & \sigma_{N2}I_T & \dots & \sigma_{NN}I_T \end{pmatrix} = \Sigma \otimes I_T, \quad \text{kur } \Sigma \text{ yra kovariacijų matrica:}$$

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1N} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2N} \\ \dots & \dots & \dots & \dots \\ \sigma_{N1} & \sigma_{N2} & \dots & \sigma_{NN} \end{pmatrix}, \quad \text{su elementu } \hat{\sigma}_{ij} = E(\varepsilon_{jt}\varepsilon_{it}), \quad \text{kuris nekinta kintant laikui.}$$

Svoriai skaičiuojami tokiu būdu: $\sigma_{ij} = \sum_t (y_{it} - \hat{y}_{it})^2 / \max(T_i, T_j)$.

Maksimumo funkcijos naudojimas yra naudingas nesubalansuotų duomenų atveju. Parametrų įverčiai ir jų kovariacijos matrica skaičiuojama naudojant apibendrintų mažiausių kvadratų metodą (Generalized Least Squares - GLS).

Norėčiau paminėti dar porą modelių, kurių, deja, nepavyko realizuoti su E-Views, tačiau manau, kad juos paminėti yra naudinga.

2.2.3. Atsitiktinių koeficientų modelis

Šis modelis tarsi praplečia atsitiktinių efektų modelį ir taria, kad ne tik laisvasis narys, bet ir krypties koeficientas kiekvienai šaliai yra atsitiktiniai kintamieji. Tuomet modelis atrodo taip:

$$Y_{it} = \sum_{k=0}^K \beta_{ik} X_{ikt} + \varepsilon_{it},$$

čia $\beta_{ik} = \bar{\beta}_k + \delta_{ik}$ ir δ_{ik} yra atsitiktinė koeficiento kiekvienai šaliai variacija apie vidurkį $\bar{\beta}_k$. Laisvasis narys laikomas vienetu. Svarbiausias klausimas vertinant šį modelį yra toks: ar teisinga prielaida apie atsitiktinę variaciją?

2.2.4. Koeficientai kaip funkcijos nuo endogeninių kintamųjų

Šiame modelyje tariama, kad krypties koeficientai priklauso nuo egzogeninių kintamųjų. Egzogeniniai kintamieji gali varijuoti laike ir tarp šalių. Čia pateikiamas modelis, kai turime variaciją tik tarp šalių, nes turime daug šalių ir trumpas laiko eilutes.

$$Y_{it} = \sum_{k=1}^{K_1} \beta_{1k} X_{1ikt} + \sum_{k=1}^{K_2} \beta_{2k} X_{2ikt} + \varepsilon_{it}.$$

Tokiame modelyje, egzogeniniai kintamieji yra išskaidomi į dvi dalis: X_{1ikt} ir X_{2ikt} . Koeficientai

β_{1k} yra fiksuotos konstantos, o $\beta_{2k} = \sum_{j=1}^J \gamma_{jk} z_{ijk} + \eta_{ik}$, kur z_{ijk} yra egzogeniniai kintamieji, γ_{jk}

yra koeficientas, o η_{ik} yra nežinoma konstanta. Sujungus šias dvi lygtis, gauname:

$$Y_{it} = \sum_{k=1}^{K_1} \beta_{1k} X_{1ikt} + \sum_{k=1}^{K_2} \sum_{j=1}^J \gamma_{jk} w_{ijkt} + \sum_{k=1}^{K_2} \eta_{ik} X_{2ikt} + \varepsilon_{it}, \quad \text{kur } w_{ijkt} = z_{ijk} x_{2ikt}.$$

3. Modeliai. Tiesinis trendas

Dažniausiai skverbties modeliai yra naudojami pardavimų priklausomybei nuo laiko ištirti. Kaip jau minėjau, yra keletas panelinių duomenų tipų: pastovios konstantos modeliai, fiksuotų efektų modeliai ir atsitiktinių efektų modeliai. Svarbios yra heteroskedastiškumo ir autokoreliacijos problemos.

3.1. Modelis su tiesiniu trendu kiekvienai šaliai

Pirmas mūsų nagrinėjamas panelinių modelių tipas turi pastovius kintamuosius: laisvuosius narius ir posvyrio koeficientus. Tuomet, kai nėra reikšmingi nei šalies, nei laiko efektai, galima sukaupti visus duomenis kartu ir naudoti paprastas regresijas, koeficientus vertinant mažiausių kvadratų metodu. Mūsų atveju vertinsime 23 lygtis, taigi, turėtume gauti 46 koeficientus. Kai kurioms šalims (Jordanijai ir Meksikai) negalime sudaryti regresinės lygties, nes turime tik po vieną stebėjimą. Šis vertinimo būdas yra tik indikatyvus, toliau pereisime prie sujungtų duomenų vertinimo (pooled estimation).

$$PC_{it} = \alpha_i + \beta_1 * TIME + \varepsilon_{it}, \quad i = \text{country}, \quad t = \text{time}$$

[vertintos lygtys:

Nr.	Šalis	C	TIME	R ²	Durbin - Watson
1	PC_AZE	-6624.848 (2421.293)	3.324000 (1.211252)	0.715127	2.034409
2	PC_CHI	-10181.63 (820.5147)	5.096429 (0.410052)	0.968647	0.939920
3	PC_CRO	-6867.860 (848.2208)	3.440000 (0.423792)	0.942766	2.215717
4	PC_DEN	-9849.055 (176.3349)	4.956044 (0.088278)	0.996207	2.009962
5	PC_EGY	-3919.110 (NA)	1.960000 (NA)	1.000000	1.999996
6	PC_EST	-8814.729 (617.2579)	4.414286 (0.308552)	0.971520	1.043289
7	PC_FRA	-9057.156 (694.8539)	4.543333 (0.347427)	0.960676	0.803091
8	PC_GER	-8157.825 (177.1736)	4.103013 (0.088653)	0.996279	1.961293
9	PC_GRE	-5675.125 (744.2586)	2.845316 (0.371943)	0.921285	0.887169
10	PC_JOR	NA	NA	NA	NA
11	PC_MEX	NA	NA	NA	NA
12	PC_NET	-8024.767	4.042857		

		(248.3740)	(0.124498)	0.987822	0.728680
13	PC_POR	-7743.411 (427.3442)	3.885000 (0.213672)	0.979265	1.364551
14	PC_ROM	-7300.000* (2253.918)	3.650000* (1.125833)	0.913125	3.000000
15	PC_RUS	-5996.000 (NA)	3.000000 (NA)	1.000000	1.999979
16	PC_SLO	-7070.047 (950.2880)	3.551429 (0.475500)	0.933092	2.587763
17	PC_SOK	-21321.16 (3931.650)	10.68393 (1.965332)	0.831234	1.671840
18	PC_SPA	-6120.154 (464.3925)	3.076923 (0.232428)	0.940939	0.771638
19	PC_SWE	-9176.155 (770.7296)	4.619048 (0.385268)	0.959931	1.395714
20	PC_TUN	-937.4800 (NA)	0.470000 (NA)	1.000000	2.000000
21	PC_TWN	-6534.856 (468.8829)	3.289156 (0.235146)	0.911487	0.132059
22	PC_UAE	-15396.97 (1336.378)	7.710000 (0.668356)	0.985193	2.440184
23	PC_USA	-8813.788 (261.4942)	4.431778 (0.130845)	0.991358	0.999718
	Vidurkis	-8265.816	4.147264		

Lentelė 3: Regresijos lygtys visoms šalims. Skliausteliuose yra standartinis nuokrypis. * žymi nereikšmingus kintamuosius.

Dažniausiai paprastos regresijos nėra geriausias koeficientų įvertinimo būdas, nes mes tariame, kad šalys yra nesusijusios viena su kita ir, kad kažkokie pokyčiai, vykstantys vienoje šalyje, neturi jokio poveikio kitose. Šios prielaidos būtų klaidingos, kadangi ekonomikos augimo lėtėjimas ar greitėjimas vienoje šalyje, stipriai įtakoja ekonominius pokyčius kitose. Reikia ieškoti geresnio modelio, kuris įtrauktų ir šalių tarpusavio sąveikas. Taip pat reikia pastebėti, kad kai kurių šalių laiko eilutės yra per trumpos, kad galėtume gauti gerus įverčius, tai yra Egiptas, Jordanija, Meksika, Rumunija, Rusija bei Tunisas. Trijų šalių: Azerbaidžano, Pietų Korėjos ir Jungtinių Arabų Emyratų laiko eilutės elgiasi labai išskirtinai ir todėl gali iškreipti įverčius; šių šalių į tolimesnę analizę neįtrauksime. Taigi toliau analizuosime 14 šalių.

3.2. Konstanta bendra visoms šalims

Norėdami įvertinti ir tarpusavio sąsajas, regresijoje naudosime sujungtus (pooled) duomenis su tokia pačia konstanta, t.y. vertinsime modelį:

$$PC_{it} = \alpha + \beta_1 * TIME + \varepsilon_{it}$$

Endogeninis kintamasis: **PC?**

Metodas: **Pooled LS / Pooled EGLS (Cross-section weights) / Pooled EGLS (Cross-section SUR).**

Data: **2007.04.10**

Periodas: **1984 2004**

Laiko stebėjimų skaičius: **21**

Šalių skaičius: **14**

Nesubalansuotų panelinių duomenų skaičius: **145**

Kintamieji	Koeficientai	Stand. paklaida	t-reikšmė	p-reikšmė
C (Pooled LS)	-4959.376	615.4500	-8.058129	0.0000
C (Weighted PLS)	-5736.561	368.2502	-15.57789	0.0000
C (SUR)	-5461.761	255.6899	-21.36088	0.0000
TIME (Pooled LS)	2.498498	0.308014	8.111631	0.0000
TIME (Weighted PLS)	2.887084	0.184369	15.65926	0.0000
TIME (SUR)	2.749802	0.127920	21.49623	0.0000
	Pooled Least Squares	Pooled Least Squares Weighted	SUR	
R ²	0.315129	0.631645 (0.611374)	0.888622	
R ² (adjusted)	0.310340	0.629069	0.887843	
RSS	38407.72	36885.79 (38915.69)	57.91698	
Durbin-Watson	0.024053	0.053944 (0.020389)	0.487839	

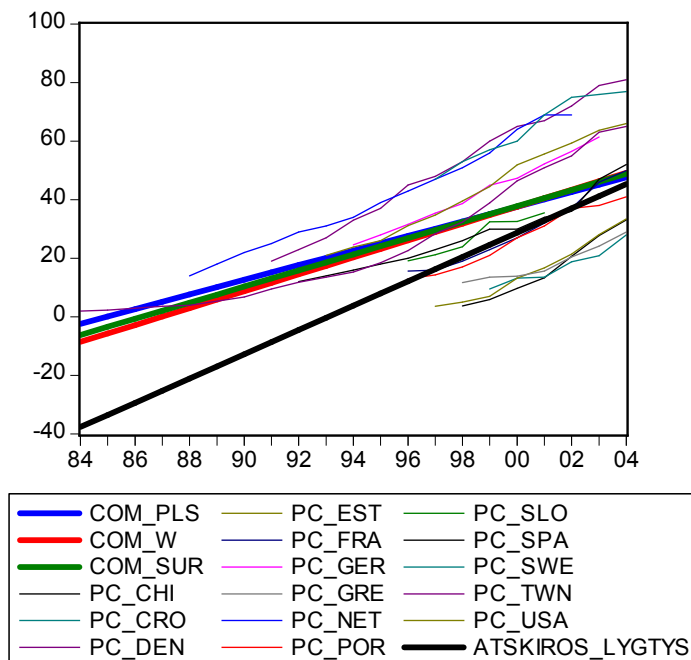
Lentelė 4: Sujungtos (pooled) regresijos lygtys su bendru (common) laisvuju nariu, skliausteliuose statistinės reikšmės be svorių (unweighted).

Lentelėje pateikti trijų skirtingų vertinimo metodų rezultatai: Pooled LS – tai MKM vertinimas be svorių. Weighted – MKM vertinimas su svoriais, jei paklaidos heteroskedastinės ir nekoreliuotos, SUR (Seemingly Unrelated Regression) vertinimas, jei paklaidos heteroskedastinės ir koreliuotos. Durbin –Watson statistika yra arti 0 visais atvejais, nors, reikia pripažinti, kad SUR vertinimo metodas ją šiek tiek pagerina, taip pat liekanų koreliacijos lentelėje, pateiktoje apačioje, matome, kad liekanos yra iš tiesų koreliuotos (koreliacija paskaičiuota modeliui įvertintam MKM be svorių). Taigi geriausiai čia tiktų vertinimas SUR metodu.

	_CHI	_CRO	_DEN	_EST	...	_SPA	_SWE	_TWN	_USA
_CHI	1.000	0.816	-0.609	0.855	...	0.387	-0.848	-0.363	-0.667
_CRO	0.816	1.000	-0.572	0.697	...	0.258	-0.793	-0.416	-0.665
_DEN	-0.609	-0.572	1.000	-0.665	...	-0.584	0.715	0.499	0.886
_EST	0.855	0.697	-0.665	1.000	...	0.472	-0.927	-0.345	-0.692
...
_SPA	0.387	0.258	-0.584	0.472	...	1.000	-0.353	0.094	-0.399
_SWE	-0.848	-0.793	0.715	-0.927	...	-0.353	1.000	0.469	0.797
_TWN	-0.363	-0.416	0.499	-0.345	...	0.094	0.469	1.000	0.598
_USA	-0.667	-0.665	0.886	-0.692	...	-0.399	0.797	0.598	1.000

Lentelė 5: Koreliacinė liekanų matrica, vertinant tiesinį modelį su vienodu laisvuju nariu, MKM be svorių.

Matome, kad vertindami visais trim metodais koeficientus gauname ganėtinai panašius. Be to, tokio modelio atveju turėtume, kad visoms šalims laisvasis narys ir koeficientas prie laiko sutampa. Iš pirmo grafiko, esančio viršuje, akivaizdžiai matome, kad laisvasis narys skiriasi. Vis dėlto pažiūrėkime kaip atrodo grafikas visiems trimis iki šiol nagrinėtiems atvejais.



Grafikas 2: Tiesiniu modelių apskaičiuotos regresijos tiesės sujungtiems (pooled) duomenims su bendru (common) laisvuju nariu.

Grafike matome išbrėžtas visų šalių kreives, o ant viršaus keturias regresijos tieses, iš kurių juoda tiesė yra gauta įvertinus lygtis kiekvienai šaliai atskirai ir imant regresijų įvertintų koeficientų vidurkį, kadangi šio darbo tema yra panelinis modelis, ši kreivė yra tikrai indikatyvi, išbrėžta palyginimui. Kitos trys tiesės (raudona, žalia ir mėlyna), gautos imant koeficientus iš sujungtų duomenų (pooled) modelio su visoms šalims nekintančiu laisvuju nariu ir vertinant jį skirtingais metodais (COM_ – common constant, PLS – Pooled Least Squares, W – Weighted Pooled Least Squares, SUR – Seemingly Unrelated regression). Vertinant sujungtus duomenis su bendru laisvuju nariu, krypties koeficientas blogai atspindi realybę ir iš tiesų yra naivus vertinimo būdas. Matome, kad šis modelis nėra tinkamas, taigi ieškokime toliau. Vertinsime fiksuotų efektų ir atsitiktinių efektų modelius. Pamėginsime juos palyginti.

3.3. Fiksuotų efektų modeliai

Kita panelinių duomenų modelių grupė vadinamieji mažiausių kvadratų fiktyvių kintamųjų modeliai (Least Squares Dummy Variable - LSDV).

Gali būti, kad turime vienodus krypties koeficientus, tačiau skirtingus laisvuosius narius kiekvienai šaliai. Mūsų atveju, skirtingi laisvieji nariai būtų skaičiuojami pagal šalis. Jei laisvieji nariai skirtųsi, tai reikštų, kad turime reikšmingą skirtumą tarp šalių išsivystymo lygio, tačiau kompiuterių įsigijimų laike dinamika būtų daugmaž vienoda. Iš grafiko matome, kad taip ir yra, krypties koeficientas atrodo esantis panašus visoms šalims. Šiuo atveju kiekvienai šaliai yra įtraukiamas fiktyvus laisvasis narys.

$$PC_{it} = \alpha + \beta_1 * TIME + \delta_i + \varepsilon_{it}$$

arba kitaip tai reiškia, kad vertiname:

$$PC_{it} = \alpha + \beta_1 * TIME + \delta_{CHI} * 1_{CHI} + \delta_{CRO} * 1_{CRO} + \delta_{DEN} * 1_{DEN} + \delta_{EST} * 1_{EST} + \\ + \delta_{FRA} * 1_{FRA} + \delta_{GER} * 1_{GER} + \delta_{GRE} * 1_{GRE} + \delta_{NET} * 1_{NET} + \delta_{POR} * 1_{POR} + \delta_{SLO} * 1_{SLO} + \\ + \delta_{SPA} * 1_{SPA} + \delta_{SWE} * 1_{SWE} + \delta_{TWN} * 1_{TWN} + \delta_{USA} * 1_{USA} + \varepsilon_{it}$$

Vertinant fiksuotų efektų modelius sujungtų duomenų mažiausių kvadratų metodu dažnos yra heteroskedastiškumo bei serijinės koreliacijos problemos. Dėl šios priežasties, kaip ir anksčiau, kartu būtų įdomu panagrinėti modelius, įvertintus naudojant svorius bei SUR (Seemingly Unrelated Regression), tačiau nesubalansuotai paneliui E-views negali vertinti SUR metodu kartu su fiksuotais efektais. Taigi, teks pasitenkinti vertinimu be svorių ir su jais.

3.4. Atsitiktinių efektų modeliai

Laisvasis narys laikomas atsitiktiniu, taigi svarbiausia yra laisvojo nario α vidurkis ir jo dispersija, o ne individualus kiekvienos šalies vertinimas. Šiuo atveju modelis tampa tokios formos, kur δ_i yra atsitiktinė šalies i variacija apie vidurkį:

$$PC_{it} = \alpha + \beta_1 * TIME + (\delta_i + \varepsilon_{it}),$$

arba

$$PC_{it} = \alpha + \beta_1 * TIME + (\delta_{CHI} * 1_{CHI} + \delta_{CRO} * 1_{CRO} + \delta_{DEN} * 1_{DEN} + \delta_{EST} * 1_{EST} + \\ + \delta_{FRA} * 1_{FRA} + \delta_{GER} * 1_{GER} + \delta_{GRE} * 1_{GRE} + \delta_{NET} * 1_{NET} + \delta_{POR} * 1_{POR} + \delta_{SLO} * 1_{SLO} + \\ + \delta_{SPA} * 1_{SPA} + \delta_{SWE} * 1_{SWE} + \delta_{TWN} * 1_{TWN} + \delta_{USA} * 1_{USA} + \varepsilon_{it})$$

Dažnai yra sunku pasirinkti tarp fiksuotų ir atsitiktinių efektų modelių. Fiksuotų efektų modelis naudoja fiktyvius kintamuosius, tam, kad išskirti šalį ir naudoja daug laisvės laipsnių. Atsitiktinių efektų modelis jų naudoja mažiau, bet taria, kad individualūs efektai nekoreliuoja su kitais kintamaisiais. Jei ši prielaida neteisinga, tuomet atsitiktinių efektų modelio įverčiai bus nesuderinti. Palyginkime rezultatus tarp fiksuotų efektų modelio vertinimų bei atsitiktinių efektų modelio.

Endogeninis kintamasis: **PC?**

Metodas: **Fixed effects: Pooled LS / Pooled EGLS (Cross-section weights). Random effects.**

Data: **2007.04.12**

Periodas: **1984 2004**

Laiko stebėjimų skaičius: **21**

Šalių skaičius: **14**

Nesubalansuotų panelinių duomenų skaičius: **145**

Kintamieji	Koeficientai	Stand. Paklaida	t-reikšmė	p-reikšmė
C (Pooled LS)	-7590.418	178.8136	-42.44878	0.0000
C (Weighted)	-8002.176	127.0539	63.24165	0.0000
TIME (Pooled LS)	3.815258	0.089491	42.63293	0.0000
TIME (Weighted)	4.021332	0.063587	63.24165	0.0000
C (Random)	-7567.615	178.7010	-42.34792	0.0000
TIME (Random)	3.801773	0.089379	42.53534	0.0000
Fiksuoti efektai	Pooled Least Squares	Pooled Least	Atsitiktiniai efektai	

		Squares Weighted	
PC_CHI	-27.59121	-28.18527	-23.15893
PC_CRO	-28.52170	-29.21880	-24.02973
PC_DEN	20.08219	20.20939	24.08502
PC_EST	-25.95644	-26.44746	-21.57737
PC_FRA	-10.58770	-10.97569	-6.366617
PC_GER	7.670828	7.591952	11.72876
PC_GRE	-25.56169	-26.15575	-21.15121
PC_NET	19.71081	20.35319	23.69065
PC_POR	-13.50992	-13.89791	-9.264365
PC_SLO	-6.892592	-6.971468	-2.711159
PC_SPA	-4.929834	-4.905674	-0.784902
PC_SWE	22.24356	21.75254	26.16897
PC_TWN	6.513690	7.362143	10.56186
PC_USA	8.743080	8.664204	12.80903
R ²	0.963242	0.98379 (0.981742)	0.922803 (0.185028)
R ² (adjusted)	0.959283	0.982021	0.922263
Durbin-Watson	0.319192	0.758104 (0.308341)	0.274335 (0.014401)
RSS	2061.418	1907.323 (2145.501)	2399.196 (45703.81)

Lentelė 6: Regresijos lygtys su fiksuotais ir atsitiktiniais efektais, skliausteliuose statistikos be svorių (unweighted).

Viršutinėje lentelės dalyje matome įvertintus laisvuosius narius, bei koeficientus prie TIME, įvertintus modeliais su fiksuotais ir atsitiktiniais efektais, tolesniuose modeliuose fiksuotų bei atsitiktinių efektų nepateiksime, nes jie visuomet svyruoja apie nulį, apsiribosime tik Hausman testu bei testu, tikrinančiu ar fiksuoti efektai nėra pertekliniai (redundant). Tiesą sakant, koeficientų reikšmes gauname labai panašias į tas, kurias gavome vertindami modelį kiekvienai šaliai atskirai.

Naudodami Hausman testą, pamėginkime pažiūrėti, kuris modelis mums labiau tinka: fiksuotų ar atsitiktinių koeficientų. Testui naudojama Chi – kvadrato statistika.

Nulinė hipotezė	Atsitiktinių efektų modelis	
Chi Sq. Statistika	Chi Sq. Laisvės laipsniai	p- reikšmė
9.100311	1	0.0026

Matome, kad nulinė hipotezė yra atmetama su pasirinktu reikšmingumo lygmeniu 0.05, reikia rinktis fiksuotų efektų modelį. Taip pat pateikiama lentelė, rodanti abiejų metodų koeficientų įverčius ir skirtumo variaciją bei p-reikšmes, kai nulinė hipotezė yra, kad skirtumas tarp šių įverčių yra nereikšmingas.

Nulinė hipotezė	TIME (Fixed) = TIME (Random)			
Kintamasis	Fixed	Random	Var (Diff)	p- reikšmė
TIME	3.815258	3.801773	0.000020	0.0026

Kaip matome iš lentelės, ši hipotezė yra atmetama, įverčiai reikšmingai skiriasi. Patikrinkime ar fiksuoti efektai nėra pertekliniai (redundant).

Fiksuotų efektų testas	Statistika	Laisvės laipsniai	p- reikšmė
Cross-section F	118.271095	(13, 129)	0.0000
Cross-section Chi-sq	371.009040	13	0.0000

Matome, kad nulinė hipotezė yra atmetama, vadinasi fiksuoti efektai nėra pertekliniai.

Dabar reikia parinkti svorius. Įvertinę modelį MKM be svorių, pažiūrime kaip atrodo lieknų kovariacinė matrica:

	_CHI	_CRO	_DEN	_EST	...	_SPA	_SWE	_TWN	_USA
_CHI	9.928	-0.908	1.933	6.024	...	3.189	1.481	2.534	0.892
_CRO	-0.908	2.506	-0.215	-0.734	...	0.178	-2.403	-0.731	-0.684
_DEN	1.933	-0.215	22.667	1.244	...	-12.177	1.986	9.831	7.379
-EST	6.024	-0.734	1.244	4.883	...	2.084	0.506	1.956	0.884
...
_SPA	3.189	0.178	-12.177	2.084	...	15.951	-2.217	-4.569	-4.813
_SWE	1.481	-2.403	1.986	0.506	...	-2.217	8.068	2.774	3.019
_TWN	2.534	-0.731	9.831	1.956	...	-4.569	2.774	48.670	6.586
_USA	0.892	-0.684	7.379	0.884	...	-4.813	3.019	6.586	6.570

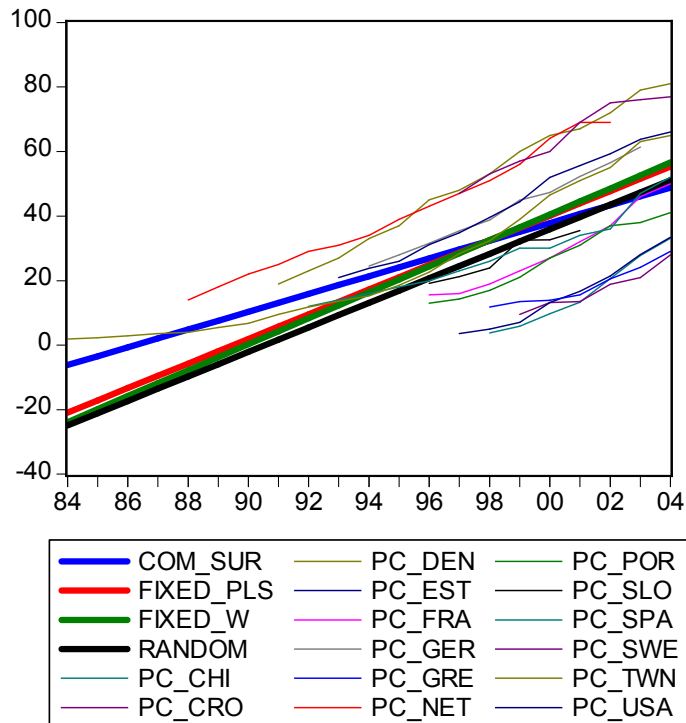
Lentelė 7: Kovariacinė liekanų matrica, vertinant tiesinio trendo modelį su fiksuotais efektais, MKM be svorių.

Matome, kad dispersija aiškiai skiriasi, taigi reikėtų rinktis MKM su svoriais. Paskaičiavę modelį su svoriais pažiūrėkime koreliacinę paklaidų matricą, paklaidos yra šiek tiek koreliuotos, tačiau SUR metodo, deja, pabandyti negalime, nes to neleidžia E-Views.

	_CHI	_CRO	_DEN	_EST	...	_SPA	_SWE	_TWN	_USA
_CHI	1,000	-0,313	0,104	0,817	...	0,197	0,033	0,082	0,010
_CRO	-0,313	1,000	-0,029	-0,318	...	0,013	-0,535	-0,066	-0,177
_DEN	0,104	-0,029	1,000	0,055	...	-0,706	0,102	0,197	0,512
-EST	0,817	-0,318	0,055	1,000	...	0,196	-0,158	0,074	0,005
...
_SPA	0,197	0,013	-0,706	0,196	...	1,000	-0,230	-0,135	-0,476
_SWE	0,033	-0,535	0,102	-0,158	...	-0,230	1,000	0,094	0,389
_TWN	0,082	-0,066	0,197	0,074	...	-0,135	0,094	1,000	0,276
_USA	0,010	-0,177	0,512	0,005	...	-0,476	0,389	0,276	1,000

Lentelė 8: Koreliacinė liekanų matrica, vertinant tiesinio trendo modelį su fiksuotais efektais, MKM su svoriais.

Taigi, nors ši modelio forma apskritai nėra tinkama, bet geriausias vertinimo būdas būtų fiksuotų efektų modelis įvertintas su svoriais. Pabraižykime grafikus, kad galėtume vizualiai pažiūrėti, kaip atrodo trendas:



Grafikas 3: Regresijos tiesės modeliams su bendra konstanta visoms šalims, fiksuotais ir atsitiktiniais efektais.

COM_SUR - modelis įvertintas su visoms šalims bendra (common) konstanta SUR metodu.
 FIXED – modelis įvertintas su fiksuotais efektais, kur PLS yra Pooled Least Squares, W – Pooled Least Squares Weighted.
 RANDOM reiškia atsitiktinių efektų modelį.
 Matome, kad vizualiai tiek fiksuotų, tiek atsitiktinių efektų modelių krypties koeficientas yra beveik identiškas, o modelio su bendra (common) konstanta krypties koeficientas blogai atspindi skverbties koeficiento dinamiką.

3.5. Modelis su tiesiniu trendu ir AR nariais

Pamėginkime į modelį įtraukti AR(1) ir AR(2) procesą. Būtų įdomu pasižiūrėti ar MA nariai būtų reikšmingi, tačiau naudojamas paketas MA narių sistemoms įtraukti neleidžia. Kadangi paketas E-Views, vertindamas atsitiktinių efektų modelį, neleidžia įtraukti autoregresijos narių, tai vertinsime tik fiksuotų efektų modelius, kadangi liekanų kovariacijos yra mažos ir šis modelis nėra vienas iš pagrindinių, tai vertinsime tik MKM be svorių ir į tolesnę analizę nesigilinsime.

$$AR(1): \quad PC_{it} = \alpha + \beta_1 * PC_{i,t-1} + \delta_i + \varepsilon_{it}$$

Endogeninis kintamasis: **PC?**

Metodas: **Pooled LS**

Data: **2007.04.16**

Periodas: **1984 2004**

Laiko stebėjimų skaičius: **20**

Šalių skaičius: **14**

Nesubalansuotų panelinių duomenų skaičius: **131**

Kintamieji	Koeficientai	Stand. Paklaida	t-reikšmė	p-reikšmė
C	-9527.166	631.9129	-15.07671	0.0000
TIME	4.782708	0.315261	15.17063	0.0000
AR(1)	0.793157	0.042571	18.63146	0.0000
R ²				0.992037
R ² (adjusted)				0.990998
RSS				395.5503
Durbin-Watson				2.058426

Lentelė 9: Fiksuotų efektų modelis su tiesiniu trendu ir AR(1) nariu.

$$AR(2): \quad PC_{it} = \alpha + \beta_1 * PC_{i,t-2} + \delta_i + \varepsilon_{it}$$

Endogeninis kintamasis: **PC?**

Metodas: **Pooled LS**

Data: **2007.04.16**

Periodas: **1984 2004**

Laiko stebėjimų skaičius: **19**

Šalių skaičius: **14**

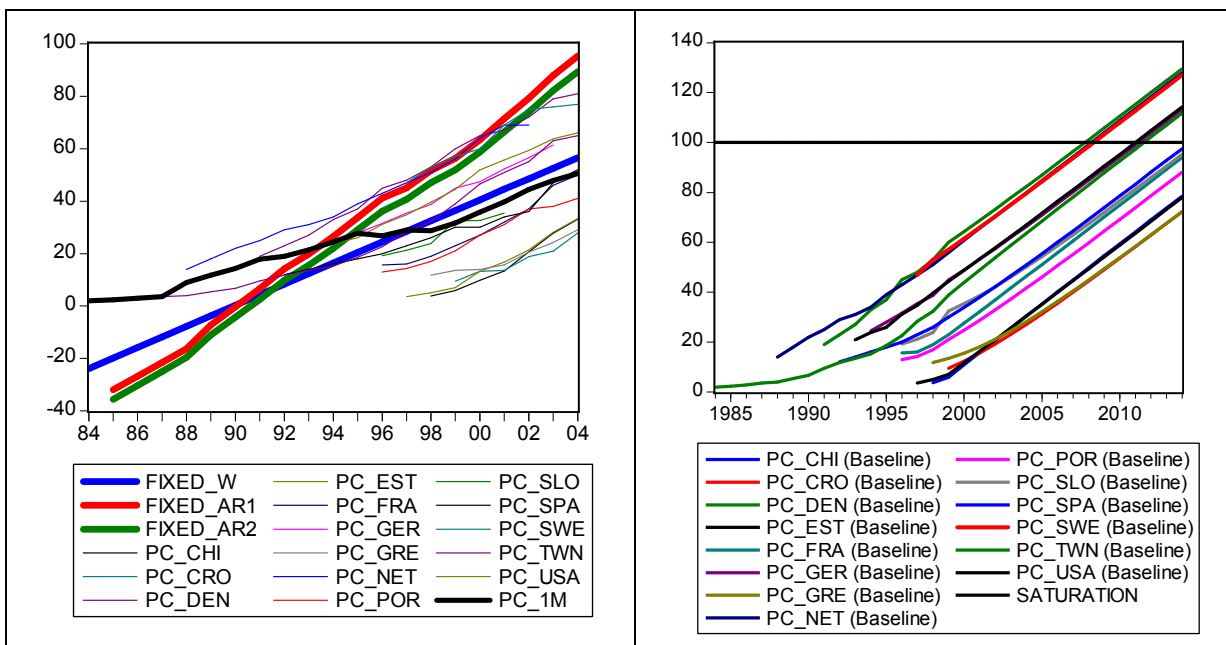
Nesubalansuotų panelinių duomenų skaičius: **117**

Kovertavimas pasiektas po 4 iteracijų

Kintamieji	Koeficientai	Stand. Paklaida	t-reikšmė	p-reikšmė
C	-9985.455	623.3115	-16.02001	0.0000
TIME	5.011860	0.310849	16.12315	0.0000
AR(2)	0.648521	0.058449	11.09541	0.0000
R ²				0.987466
R ² (adjusted)				0.985605
RSS				541.4089
Durbin-Watson				1.162066

Lentelė 10: Fiksuotų efektų modelis su tiesiniu trendu ir AR(2) nariu.

Įtraukus ir AR(1) ir AR(2), antrasis tampa nereikšmingas. Pažiūrėjime aproksimacinį grafiką ir prognozes, kurios nubrėžtos tik įtraukus narį AR(1), nes šio modelio statistikos yra geresnės.



Grafikas 4: Trendas su AR(1), AR(2) ir fiksuotų efektų su svoriais įvertinimais. Prognozė tik su AR(1).

FIXED_ – fiksuoti efektai, kur W – Pooled Least Squares Weighted, AR1 – įtraukus narį AR(1), AR2 – įtraukus narį AR(2).

PC_1M – skverbties koeficiento PC vidurkinės reikšmės, jos buvo įtrauktos skaičiuojant aproksimacines kreives.

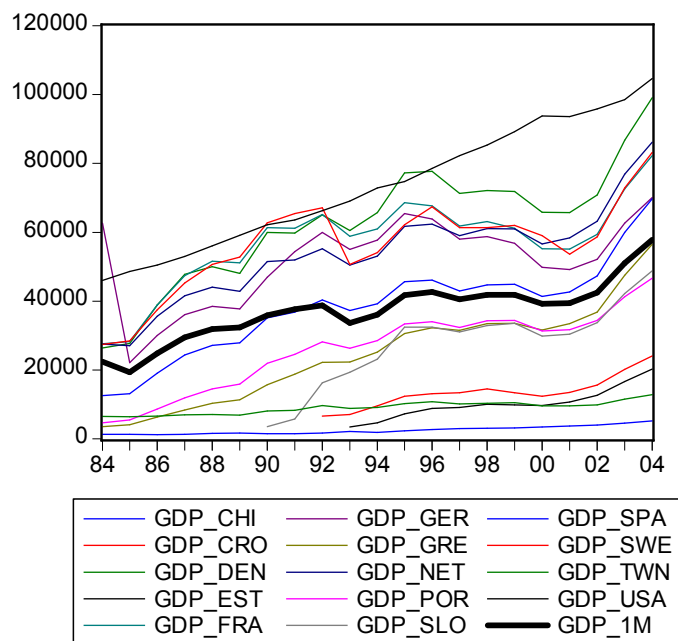
Matome, kad aproksimacija visai netinkama ir prognozė staigiai auga į viršų – prognozuojamas pastovus augimas.

3.6. Modelis su tiesiniu trendu ir egzogeniniais kintamaisiais

Pamėginau įdėti keletą papildomų egzogeninių kintamųjų. Straipsniuose, kuriuose tiriami interneto bei mobiliųjų paslaugų (telefonų, programinių paketų) skverbties koeficientai, mėginama įtraukti egzogeninius kintamuosius, taigi iš jų buvo atrinkti keli, kurie galbūt galėtų paaiškinti ir kompiuterių skverbtį. Šiek tiek apie juos.

Bendrasis vidaus produktas pagal namų ūkius. GDP.

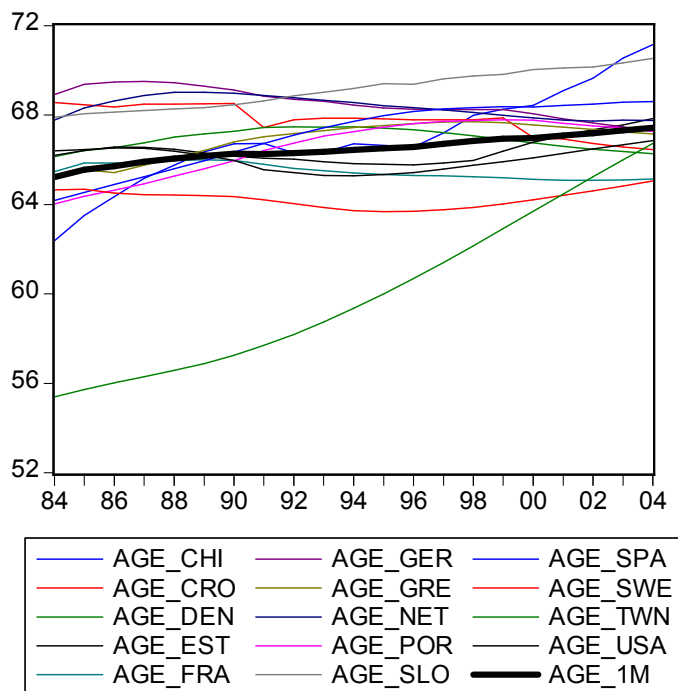
BVP rodo bendrą šalies ūkio augimą. Augant ekonomikai, turėtų augti ir kompiuterių įsigijimai. Pasirinkome BVP pagal namų ūkius todėl, kad mūsų analizuojamos šalys yra labai skirtingos. Gali būti, kad mažiau išsivysčiusiose šalyse yra didesni namų ūkiai, nei Europos ar Šiaurės Amerikos šalyse, taigi tikėtina, kad mažiau išsivysčiusiose šalyse kompiuteris perkamas vienas dideliame namų ūkiui, taigi bent jau asmeniniam vartojimui jų turėtų būti nupirka mažiau.



Grafikas 5: Bendrasis šalių produktas.

Darbingo amžiaus žmonių procentas populiacijoje. AGE.

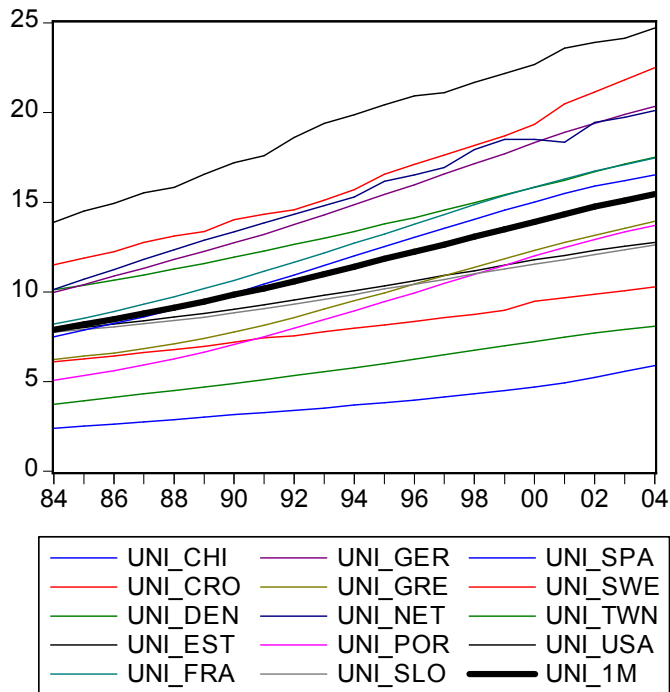
Kompiuteriais dažniausiai naudojasi darbingo amžiaus žmonės, tikėtina, kad šio skaičiaus kitimas turėtų įtakoti ir kompiuterių įsigijimus. Nors, pažiūrėjus į grafiką matome, kad daugelyje valstybių nagrinėjamas laikotarpis yra per trumpas, kad matytųsi akivaizdus kitimas.



Grafikas 6: Darbingo amžiaus žmonių procentas populiacijoje.

Universitetinį išsilavinimą turinčių žmonių procentas populiacijoje. UNI.

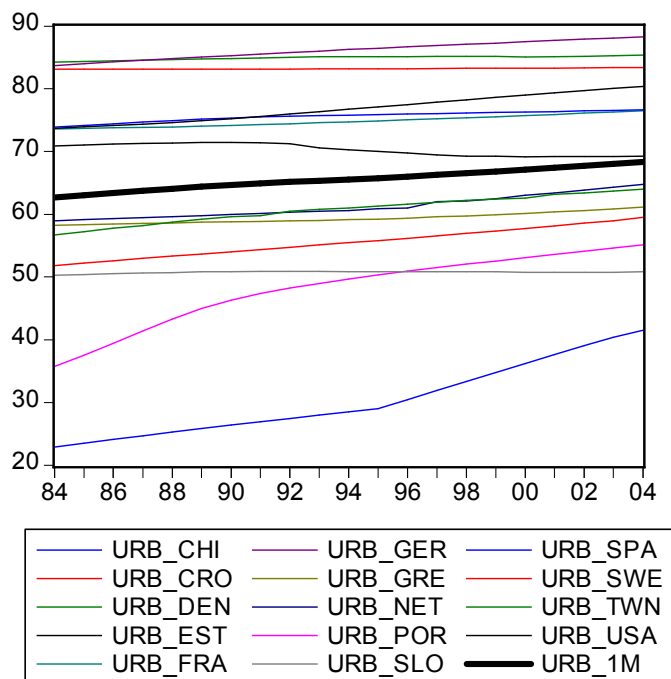
Kuo labiau išsilavinę žmonės, tuo daugiau savo veikloje naudoja kompiuterius, jei populiacija neišsilavinusi, gali būti, kad žmonės tiesiog neturės įgūdžių naudotis kompiuteriais, nemanys, kad jis naudingas, taigi pirks mažiau. Duomenys iš tiesų turi tendenciją kilti, vadinasi išsilavinusių žmonių procentas kiekvienais metais didėja.



Grafikas 7: Universitetinį išsilavinimą turinčių žmonių procentas populiacijoje.

Urbanizacijos lygis šalyse. URB.

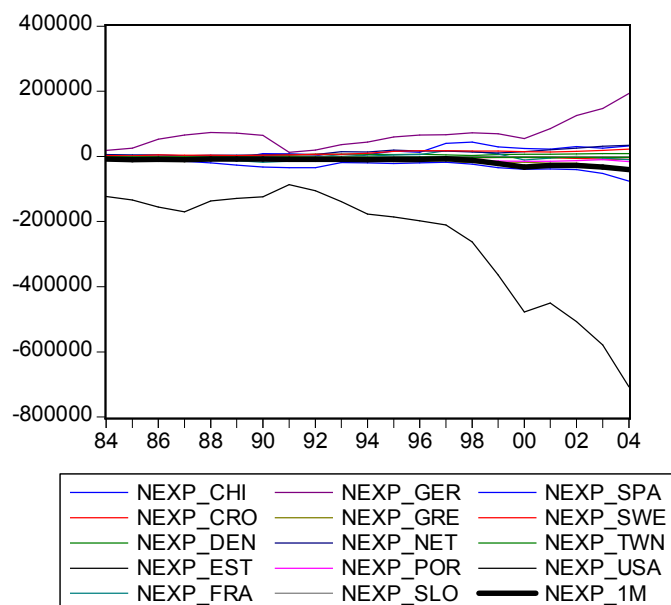
Manoma, kad pagrindinis kompiuterių vartotojas yra miestas, taigi tikėtina, kad šis kintamasis, taip pat turėtų įtakoti kompiuterių skverbties koeficientą šalyse. Tačiau miestai auga sąlyginai lėčiau ir iš grafiko matyti, kad šis kintamasis gali būti nereikšmingas.



Grafikas 8: Urbanizacijos lygis šalyse.

Grynasis eksportas. NEXP.

Grynasis eksportas galėtų atspindėti užsienio prekybos įtaką skerbties koeficientui. Tačiau kaip matome iš grafiko, šie duomenys nėra informatyvūs. Labai išsiskiria Jungtinių Amerikos Valstijų ir Vokietijos kreivės.



Grafikas 9: Grynasis eksportas.

Pasirodė esąs reikšmingas tik universitetinį išsilavinimą turinčių žmonių procentas populiacijoje (UNI). Apačioje lentelėje matome, kad papildomų egzogeninių kintamųjų įtraukimas modelio statistinių charakteristikų ženkliai nepagerino. Taigi vertinamas modelis:

$$PC_{it} = \alpha + \beta_1 * TIME + \beta_2 * UNI_{it} + \delta_i + \varepsilon_{it}$$

Įvertinsime modelį atsitiktinių efektų metodu ir su Hausman testu rinksimės modelį (fiksuotų arba atsitiktinių efektų).

Nulinė hipotezė	Atsitiktinių efektų modelis	
Chi Sq. Statistika	Chi Sq. Laisvės laipsniai	p- reikšmė
15.445920	2	0.0004

Matome, kad nulinė hipotezė yra atmetama – reikia rinktis fiksuotų efektų modelį. Taip pat pateikiama lentelė, rodanti abiejų metodų koeficientų įverčius, skirtumo variaciją ir p-reikšmes, kai nulinė hipotezė yra, kad skirtumas tarp šių įverčių yra nereikšmingas.

Nulinė hipotezė	TIME (Fixed) = TIME (Random)			
Kintamasis	Fixed	Random	Var (Diff)	P- reikšmė
TIME	2.990039	2.874921	0.028976	0.4989
UNI?	2.216487	2.459729	0.207100	0.5930

Taigi matome, nors Hausman testas teigia, kad geresnis modelis yra fiksuotų efektų, tačiau žymaus skirtumo tarp koeficientų įvertinimo nėra. Pažiūrėkime, ar fiksuoti efektai nėra pertekliniai.

Fiksuotų efektų testas	Statistika	Laisvės laipsniai	p- reikšmė
Cross-section F	118.271095	(13, 129)	0.0000
Cross-section Chi-sq	371.009040	13	0.0000

Matome, kad nulinė hipotezė yra atmetama, vadinasi fiksuoti efektai nėra pertekliniai. Pažiūrėkime kovariacijų matricą:

	CHI	CRO	DEN	EST	...	SPA	SWE	TWN	USA
CHI	12.120	0.406	1.966	8.157	...	3.014	-2.274	3.816	0.516
CRO	0.406	2.317	0.002	0.094	...	0.764	-1.697	-0.432	-0.439
DEN	1.966	0.002	19.982	1.587	...	-14.857	-0.052	12.450	4.093
EST	8.157	0.094	1.587	6.826	...	2.107	-2.234	3.291	0.740
...
SPA	3.014	0.764	-14.857	2.107	...	22.840	-2.738	-8.823	-3.755
SWE	-2.274	-1.697	-0.052	-2.234	...	-2.738	4.180	0.150	0.564
TWN	3.816	-0.432	12.450	3.291	...	-8.823	0.150	38.577	5.038
USA	0.516	-0.439	4.093	0.740	...	-3.755	0.564	5.038	3.384

Lentelė 11: Kovariacinė liekanų matrica, vertinant tiesinį fiksuotų efektų modelį su egzogeniniais kintamaisiais, MKM be svorių.

Matome, kad liekanos nėra homoskedastinės, vadinasi vertinsim fiksuotų efektų modelį su svoriais. Kadangi šis modelis taip pat nėra vienas iš svarbiausių analizuojamų, toliau į analizę nesigilinsime.

Endogeninis kintamasis: **PC?**
 Metodas: **Pooled Least Squares**
 Data: **2007.04.20**
 Periodas: **1984 2004**
 Laiko stebėjimų skaičius: **21**

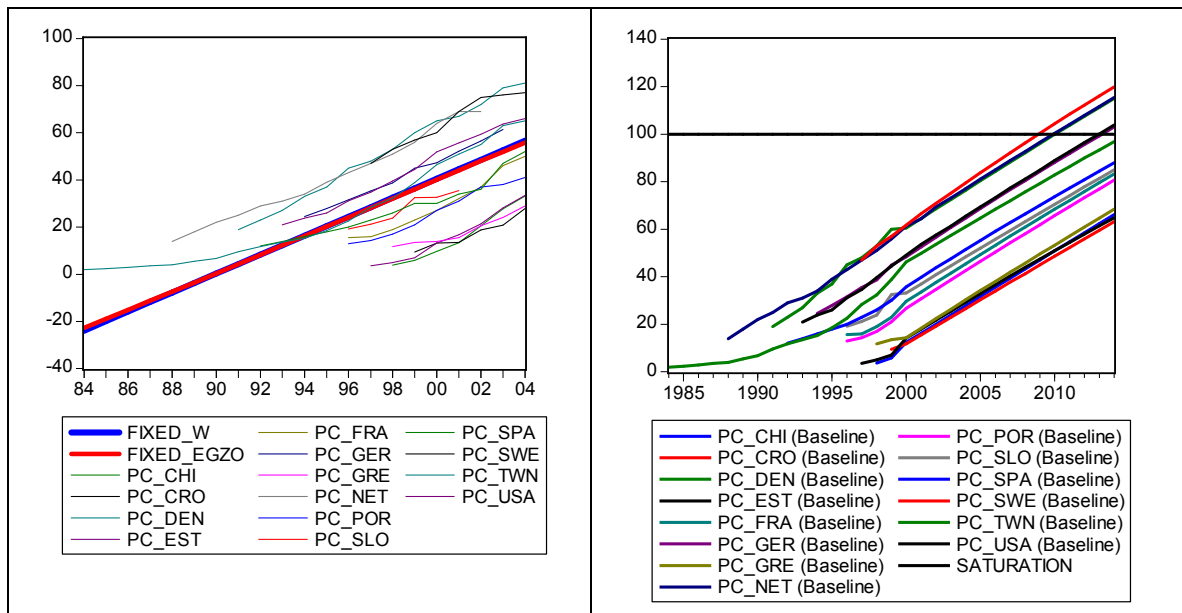
Šalių skaičius: 14

Nesubalansuotų panelinių duomenų skaičius: 145

Kintamieji	Koeficientai	Stand. Paklaida	t-reikšmė	p-reikšmė
C	-6881.682	499.5639	-13.77538	0.0000
TIME	3.452074	0.253305	13.62814	0.0000
UNI	1.264950	0.504240	2.508628	0.0134
R ²			0.966685	
R ² (adjusted)			0.962811	
RSS			1868.305	
Durbin-Watson			0.369747	

Lentelė 12: Fiksuotų efektų modelis su tiesiniu trendu ir egzogeniniais kintamaisiais, MKM su svoriais.

Pažiūrėkime grafiką ir modelio prognozę dešimčiai metų į priekį. Matome, kad didelių pasikeitimų UNI kintamasis neįneša.



Grafikas 10: Tiesinis trendas su egzogeniniais kintamaisiais ir modelio prognozė.

Matome, kad įtraukti egzogeniniai kintamieji pagerina modelio statistikas, sumažina liekanų koreliaciją, tačiau Durbin–Watson statistika lieka arti nulio. Mėginau įtraukti AR narį, tačiau tuomet egzogeninis kintamasis UNI tampa nereikšmingas.

4. Inovacijų difuzijos modeliai

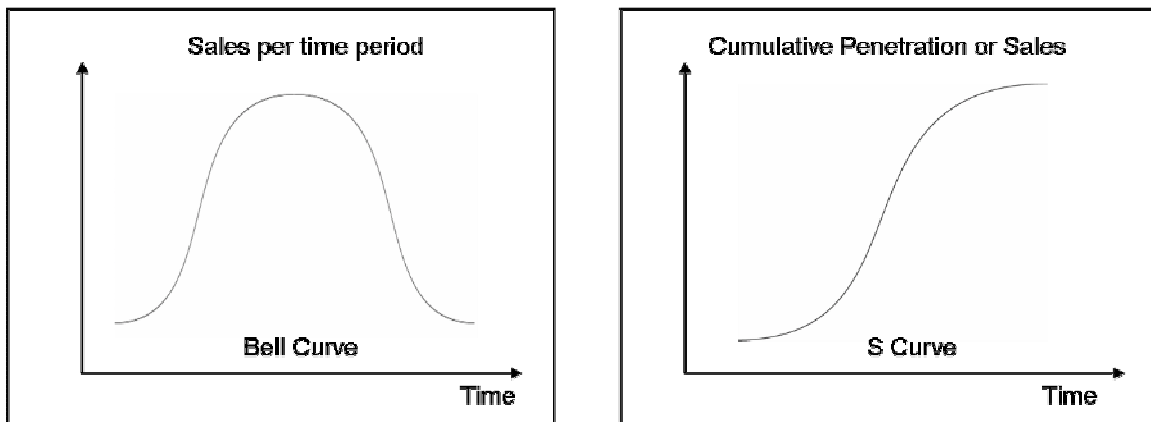
Inovacijų difuzija – tai tyrimas kaip, kodėl ir koku greičiu naujos idėjos ir technologijos plinta skirtingose kultūrose ir šalyse. Inovacijų difuzijos teorija buvo suformuota 1962 metais Everett Rogers knygoje „Inovacijų difuzija“. Jis teigė, kad visi rinkos dalyviai gali būti skirstomi į kelias grupes:

- Novatoriai (2.5%) – azartiški, išsilavinę, turintys daug prieigos prie naujos informacijos, linkę rizikuoti.
- Ankstyvi pirkėjai (13.5%) – visuomenės lyderiai, populiarūs, išsilavinę.
- Ankstyvoji dauguma (34%) – sąmoningi žmonės, gyvenantys aktyvų socialinį gyvenimą.
- Vėlyvoji dauguma (34%) – skeptikai, konservatyvūs, gyvenantis mažiau aktyvų socialinį gyvenimą.
- Atsiliekantys (16%) – žmonės bijantys rizikuoti, pasyvūs, pagrindinę informaciją gaunantys iš draugų ar kaimynų.

Taigi, kiekvieno rinkos dalyvio noras ir galimybės įsigyti naujovę priklauso nuo jų atsargumo, naujovės vertinimo, bandymo ir įsisavinimo.

Rogers teigė, kad inovacijos turėtų sklirti S-formos kreive, kadangi, ankstyvieji pirkėjai pirmieji įsigyja naujovę, tuomet jų pavyzdžiu seka didžioji dauguma žmonių, kol nauja technologija tampa įprasta rinkoje. Šis modelis sulaukė daug kritikos, buvo manoma, kad jis yra pernelyg paprastas sudėtingos realybės vaizdavimas. Tačiau S-formos difuzijos modeliavimas plačiai pasklido. Mėginsime S-formos kreives modeliuoti ir mes. Kadangi bendras laisvasis narys iš tiesų neturi prasmės, mėginsime vertinti modelius su fiksuotais ir atsitiktiniais efektais.

Pamėginkime vertinti modelį su netiesiniu trendu. Prielaida tokiam modeliui paprasta: manoma, kad visų inovacijų rinka turėtų elgtis kaip difuzija, vadinasi kažkada kreivė turėtų užsilenkti, t.y. pasiekti prisisotinimo tašką. Jei modeliuojam tiesiog tiesine regresija, darom prielaidą, kad kompiuterių pirkimas pastoviai vienodai didės, kas neturėtų būti teisinga. Manoma, kad inovacijų pardavimo kitimas laike turėtų atrodyti maždaug taip:



Grafikas 11: Varpo pavidalo pardavimų kreivė (kairėje) ir jos integralas (S - formos kreivė, dešinėje).

4.1. Kvadratinio trendo išskyrimas

Bandykime pritaikyti S pavidalo kreives, o tiksliau normaliojo pasiskirstymo Φ funkciją.

$$y_{i,t} = A_i \Phi\left(\frac{t - \mu_i}{\sigma_i}\right) = A_i \Phi_{\mu_i, \sigma_i}(t)$$

$$y_{i,t} - y_{i,t-1} = A_i \left(\Phi_{\mu_i, \sigma_i}(t) - \Phi_{\mu_i, \sigma_i}(t-1) \right) \approx A_i \exp\left(-\frac{(t - \mu_i)^2}{\sigma_i^2}\right)$$

$$\log(y_{i,t} - y_{i,t-1}) = \log A_i - \frac{(t - \mu_i)^2}{\sigma_i^2}$$

$$\log(y_{i,t} - y_{i,t-1}) = C_i(1) + C_i(2) \cdot t + C_i(3) \cdot t^2 + \varepsilon_t$$

Taigi, vertinsime modelį su kvadratinio trendu:

$$\log(PC_{i,t} - PC_{i,t-1}) = \alpha + \beta_1 * TIME + \beta_2 * TIME^2 + \delta_i + \varepsilon_t$$

Endogeninis kintamasis: **LOG(PC?-PC?(-1))**

Metodas: **Fixed effects: Pooled LS / Pooled EGLS (Cross-section weights). Random effects.**

Data: **2007.04.24**

Periodas: **1985 2004**

Ištraukta laiko stebėjimų: **20**

Šalių skaičius: **14**

Nesubalansuotų panelinių duomenų skaičius: **129**

Kintamieji	Koeficientai	Stand. paklaida	t-reikšmė	p-reikšmė
C (Pooled LS)	-27678.50	10830.26	-2.555664	0.0119
C (Weighted)	-31116.49	7411.381	-4.198475	0.0001
TIME (Pooled LS)	27.63987	10.84846	2.547816	0.0122
TIME (Weighted)	31.08731	7.424501	4.187124	0.0001
TIME ² (Pooled LS)	-0.006900	0.002717	-2.539869	0.0124
TIME ² (Weighted)	-0.007764	0.001859	-4.175632	0.0001
C (Random)	-28411.91	10691.59	-2.657408	0.0089
TIME (Random)	28.39194	10.70982	2.651019	0.0091
TIME ² (Random)	-0.007093	0.002682	-2.644531	0.0092
Fiksuoti efektai	Pooled Least Squares	Pooled Least Squares Weighted	Atsitiktiniai efektai	
R ²	0.323722	0.493298 (0.491268)	0.190641 (0.170883)	
R ² (adjusted)	0.233951	0.426037	0.177794	
RSS	49.90784	49.77343 (49.97290)	56.93887 (61.18709)	
Durbin-Watson	2.123951	1.919322 (2.121388)	1.862902 (1.733560)	

Lentelė 13: Regresijos lygtys su kvadratinio trendu.

Matome, kad Durbin – Watson statistika yra arčiau 2, vadinasi serijinės koreliacijos problemos neturėtų būti.

Šiuo atveju Hausman testas veikia blogai, matyt E-Views nesupranta, kad skaičiuojant testą, dešinėje regresijos lygties pusėje kintamasis TIME^2 turi būti įvertintas atskirai, taigi šiuo atveju, parenkant geresnį vertinimo būdą pasikliauju nuojauta. Be to, iš tiesų iš lentelės matome, kad koeficientų įvertinimai mažai skiriasi.

Pažiūrėkime ar fiksuoti efektai nėra pertekliniai:

Fiksuotų efektų testas	Statistika	Laisvės laipsniai	p- reikšmė
Cross-section F	1.909262	(13, 113)	0.0361
Cross-section Chi-sq	25.614708	13	0.0191

Matome, kad šiuo atveju p-reikšmė nėra akivaizdžiai lygi nuliui, tačiau, su reikšmingumo lygmeniu 0.05, hipotezę atmetame – fiksuoti efektai nėra pertekliniai.

Pažiūrėkime liekanų kovariacijos matricą.

	_CHI	_CRO	_DEN	_EST	...	_TUN	_TWN	_UAE	_USA
CHI	0.159	0.142	-0.006	0.100	...	0.008	-0.121	-0.016	-0.033
CRO	0.142	1.255	0.068	0.114	...	-0.047	-0.392	-0.037	-0.025
DEN	-0.006	0.068	0.275	-0.029	...	-0.022	0.004	0.037	0.098
EST	0.100	0.114	-0.029	0.252	...	0.047	-0.262	0.002	-0.024
...
TUN	0.008	-0.047	-0.022	0.047	...	0.163	-0.178	0.022	-0.014
TWN	-0.121	-0.392	0.004	-0.262	...	-0.178	0.702	0.016	0.068
UAE	-0.016	-0.037	0.037	0.002	...	0.022	0.016	0.165	0.050
_USA	-0.033	-0.025	0.098	-0.024	...	-0.014	0.068	0.050	0.115

Lentelė 14: Kovariacinė liekanų matrica, vertinant fiksuotų efektų modelį su kvadratiniais trendais, MKM be svorių.

Matome, kad reikšmės nėra labai didelės, tačiau pasirinkime MKM su svoriais. Pažiūrėkime liekanų koreliacijų lentelę:

	_CHI	_CRO	_DEN	_EST	...	_TUN	_TWN	_UAE	_USA
CHI	1.000	0.326	-0.024	0.508	...	0.076	-0.370	-0.095	-0.242
CRO	0.326	1.000	0.124	0.206	...	-0.088	-0.425	-0.081	-0.065
DEN	-0.024	0.124	1.000	-0.105	...	-0.118	-0.016	0.152	0.528
EST	0.508	0.206	-0.105	1.000	...	0.253	-0.629	0.015	-0.134
...
TUN	0.076	-0.088	-0.118	0.253	...	1.000	-0.537	0.145	-0.096
TWN	-0.370	-0.425	-0.016	-0.629	...	-0.537	1.000	0.030	0.208
UAE	-0.095	-0.081	0.152	0.015	...	0.145	0.030	1.000	0.355
_USA	-0.242	-0.065	0.528	-0.134	...	-0.096	0.208	0.355	1.000

Lentelė 15: Koreliacinė liekanų matrica, vertinant fiksuotų efektų modelį su kvadratiniais trendais, MKM su svoriais.

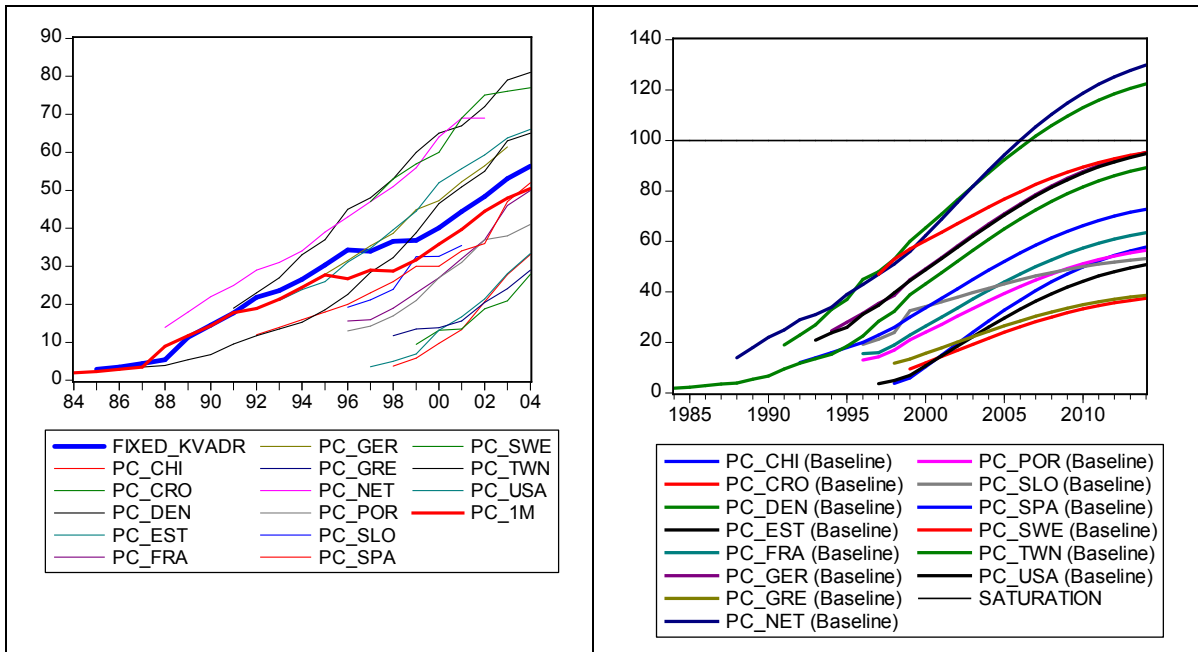
Aproksimuojanti kreivė buvo išskaičiuota formulėje vietoje $PC_{i,t-1}$ įstatant vidurkines reikšmes, grafike išbrėšime ir ją, nes iš tiesų PC reikšmės stipriai įtakoja aproksimacinę kreivę.

$$\log(PC_{i,t} - PC_{i,t-1}) = \hat{\alpha} + \hat{\beta}_1 * TIME + \hat{\beta}_2 * TIME^2$$

$$PC_{i,t} - PC_{i,t-1} = \exp(\hat{\alpha} + \hat{\beta}_1 * TIME + \hat{\beta}_2 * TIME^2)$$

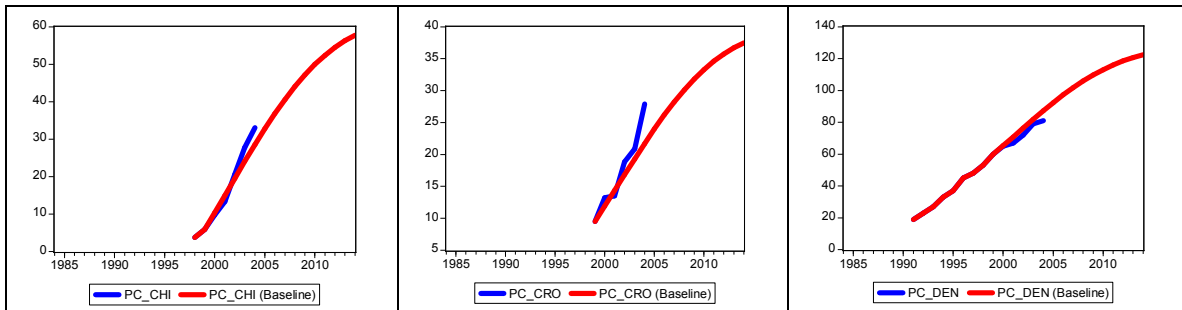
$$PC_{i,t} = \exp(\hat{\alpha} + \hat{\beta}_1 * TIME + \hat{\beta}_2 * TIME^2) + PC_{i,t-1}$$

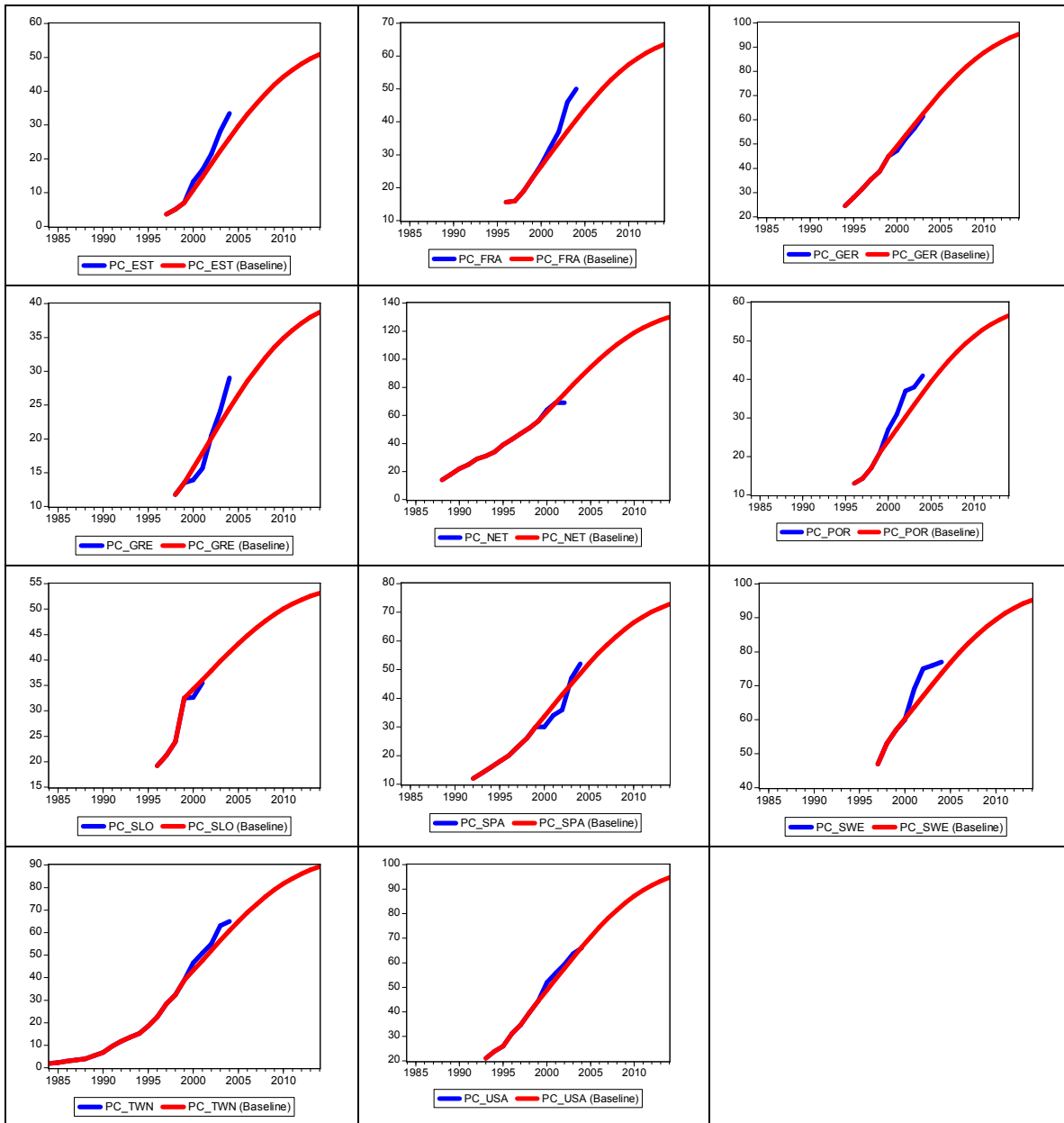
Taip pat šalia nubrėškime ir prognozes 10 metų į priekį. Iš pirmo žvilgsnio tokia prognozė lyg ir atitiktų lūkesčius, grafike apačioje matome, kad kreivės lyg ir artėja prie prisotinimo taško, tačiau kai kurioms šalims jis tampa labai aukštas, artėja prie 140, bet to, difuzijos procesą aproksimuojant kvadratinu trendu, tikėtina, kad tokia kreivė turėtų užsilenkti žemyn, kaip parabolė, tačiau pasitikrinimui pamėginau prognozuoti iki 2020 metų ir iš tiesų yra artėjama link įsisotinimo taško, o vėliau kreivė beveik pereina į tiesę.



Grafikas 12: Kvadratinis trendas bei modelio prognozės.

Pažiūrėkime, kaip atrodytų kiekvienai šaliai išbrėžus atskirai jos tikrąją PC kreivę ir prognozę viename grafike. Gali būti, kad bendra prognozė atrodo graži, tačiau gali neatitikti savo šalies realių duomenų.



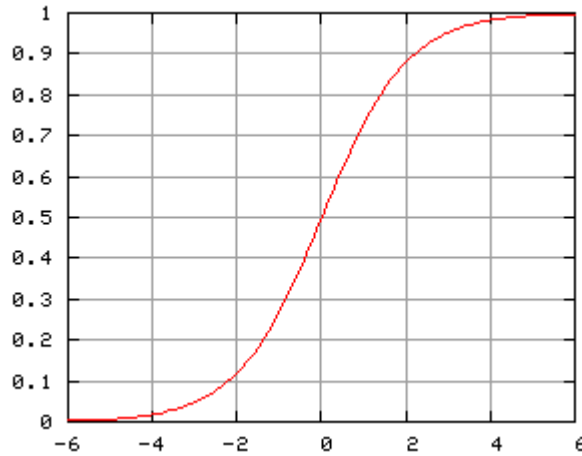


Grafikas 13: Kvadratinio trendo modelio prognozės kiekvienai šaliai.

Matome, kad didžiausias augimas prognozuojamas Danijoje ir Olandijoje, mažiausias – Graikijoje ir Kroatijoje. Tačiau aštuoniuose grafikuose iš keturiolikos matome, kad tikroji kreivė eina aukščiau prognozuojamosios, tai galėtų reikšti, kad mūsų modelis įvertina mažesnę skverbties koeficiento augimą nei iš tikrųjų bus. Šiame modelyje nėra augimo lubų. Tolesniame modelyje tarsime, kad augimo lubos yra 100%, kas reikštų, kad 100% namų ūkių turės kompiuterius.

4.2. Logistinis modelis

Pamėginsime duomenis modeliuoti logistine funkcija. Ji modeliuoja S-formos kreives. Naudojama modeliuoti tokiems duomenims, kurie turi tendencijas pradinėje stadijoje augti eksponentiškai, tuomet, kai išauga konkurencija arba, rinkta prisitotina, augimas mažėja, kol galiausiai sustoja. Reikia pastebėti, kad mūsų atveju reikšmės kis režiuose nuo 0 iki 100. Taip pat norėčiau iš karto pastebėti vieną galimą modelio trūkumą: jis yra simetrinis.



Grafikas 14: Logistinė funkcija.

Šiam modeliui vertinti pasirinksiu lubas lygias 100 ir vertinsiu tokį modelį su papildomu egzogeniniu kintamuoju UNI.

$$\log \frac{PC_{it}}{100 - PC_{it}} = \alpha + \beta_1 * TIME + \beta_2 * UNI + \delta_i + \varepsilon_{it}$$

Tačiau E-Views negali vertinti modelio, jei kairėje pusėje endogeninis kintamasis kartojasi du kartus, dėl to susikuriame naują endogeninį kintamąjį $\log ist = \log \frac{PC_{it}}{100 - PC_{it}}$ ir vertinsiu:

Endogeninis kintamasis: **LOGIST?**

Metodas: **Fixed effects: Pooled LS / Pooled EGLS (Cross-section weights). Random effects.**

Data: **2007.04.30**

Periodas: **1984 2004**

Įtraukta laiko stebėjimų: **21**

Šalių skaičius: **14**

Nesubalansuotų panelinių duomenų skaičius: **149**

Kintamieji	Koeficientai	Stand. Paklaida	t-reikšmė	p-reikšmė
C (Pooled LS)	-569.0053	20.14967	-28.23895	0.0000
C (Weighted)	-550.4375	10.74947	-51.20600	0.0000
TIME (Pooled LS)	0.285562	0.010246	27.87184	0.0000
TIME (Weighted)	0.276193	0.005456	50.62093	0.0000
UNI (Pooled LS)	-0.184676	0.025737	-7.175372	0.0000
UNI (Weighted)	-0.173191	0.011941	-14.50339	0.0000
C (Random)	-433.2170	8.535529	-50.75455	0.0000

TIME (Random)	0.216262	0.004269	50.66068	0.0000
UNI (Random)	Nereikšmingas			
Fiksuoti efektai	Pooled Least Squares	Pooled Least Squares Weighted	Atsitiktiniai efektai	
R ²	0.980467	0.994398 (0.993951)	0.944798 (0.268218)	
R ² (adjusted)	0.978195	0.993746	0.944412	
RSS	3.358944	3.163372 (3.415503)	5.424535 (125.8362)	
Durbin-Watson	0.512928	1.149645 (0.508222)	0.340100 (0.014661)	

Lentelė 16: Logistinis modelis su trendu ir egzogeniniu kintamuoju UNI, skliausteliuose statistinės reikšmės be svorių.

Taikykite Hausman testą:

Nulinė hipotezė	Atsitiktinių efektų modelis	
Chi Sq. Statistika	Chi Sq. Laisvės laipsniai	p- reikšmė
8.483875	1	0.0036

Matome, kad nulinė hipotezė yra atmetama – reikia rinktis fiksuotų efektų modelį. Taip pat pateikiama lentelė, rodanti abiejų metodų koeficientų įverčius, skirtumo variaciją ir p-reikšmes, kai nulinė hipotezė yra, kad skirtumas tarp šių įverčių yra statistiškai reikšmingas.

Nulinė hipotezė	TIME (Fixed) = TIME (Random)			
Kintamasis	Fixed	Random	Var (Diff)	P- reikšmė
TIME	0.216806	0.216262	0.000000	0.0036

Matome, kad skirtumas nėra reikšmingas. Taip pat matome, kad fiksuoti efektai nėra pertekliniai:

Fiksuotų efektų testas	Statistika	Laisvės laipsniai	p- reikšmė
Cross–section F	172.618965	(13, 129)	0.0000
Cross–section Chi-sq	422.257032	13	0.0000

Pirmiausia pažiūrėkime liekanų kovariacijų lentelę, kad žinotume ar liekanos yra homoskedastinės:

	_CHI	_CRO	_DEN	_EST	...	_SPA	_SWE	_TWN	_USA
_CHI	0.157	0.003	0.005	0.103	...	-0.013	0.041	-0.003	-0.002
_CRO	0.003	0.007	0.000	0.003	...	-0.001	-0.003	0.000	0.000
_DEN	0.005	0.000	0.005	0.005	...	-0.003	0.000	-0.001	-0.001
_EST	0.103	0.003	0.005	0.129	...	-0.020	0.043	-0.002	0.005
...
_SPA	-0.013	-0.001	-0.003	-0.020	...	0.025	-0.008	0.002	-0.001
_SWE	0.041	-0.003	0.000	0.043	...	-0.008	0.027	-0.002	0.002
_TWN	-0.003	0.000	-0.001	-0.002	...	0.002	-0.002	0.008	0.001
_USA	-0.002	0.000	-0.001	0.005	...	-0.001	0.002	0.001	0.004

Lentelė 17: Kovariacinė liekanų matrica, vertinant logistinį modelį su egzogeniniais kintamaisiais, MKM be svorių.

Nors skirtumai nėra labai dideli, tačiau pasirinkime MKM su svoriais ir pažiūrėkime liekanų koreliacijų lentelę:

	CHI	CRO	DEN	EST	...	SPA	SWE	TWN	USA
CHI	1.000	0.175	0.220	0.723	...	-0.196	0.635	-0.056	-0.023
CRO	0.175	1.000	0.042	0.165	...	-0.121	-0.155	-0.065	-0.086
DEN	0.220	0.042	1.000	0.247	...	-0.384	0.058	-0.139	-0.138
EST	0.723	0.165	0.247	1.000	...	-0.339	0.737	-0.008	0.299
...
SPA	-0.196	-0.121	-0.384	-0.339	...	1.000	-0.323	-0.032	-0.353
SWE	0.635	-0.155	0.058	0.737	...	-0.323	1.000	-0.081	0.232
TWN	-0.056	-0.065	-0.139	-0.008	...	-0.032	-0.081	1.000	0.265
USA	-0.023	-0.086	-0.138	0.299	...	-0.353	0.232	0.265	1.000

Lentelė 18: Koreliacinė liekanų matrica, vertinant logistinį modelį su egzogeniniais kintamaisiais, MKM su svoriais.

Norėdami suskaičiuoti tikrąsias reikšmes, skaičiuojame atgal:

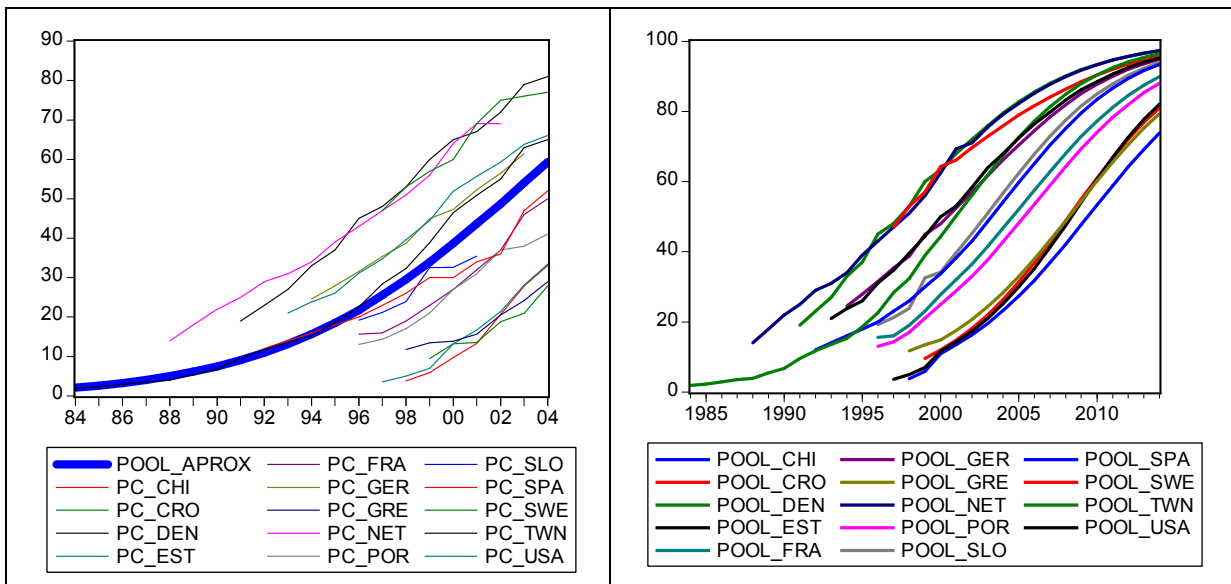
$$\frac{PC_{it}}{100 - PC_{it}} = \exp(\hat{\alpha} + \hat{\beta}_1 * TIME + \hat{\beta}_2 * UNI)$$

$$PC_{it} = (100 - PC_{it}) * \exp(\hat{\alpha} + \hat{\beta}_1 * TIME + \hat{\beta}_2 * UNI) =$$

$$= 100 * \exp(\hat{\alpha} + \hat{\beta}_1 * TIME + \hat{\beta}_2 * UNI) - PC_{it} * \exp(\hat{\alpha} + \hat{\beta}_1 * TIME + \hat{\beta}_2 * UNI)$$

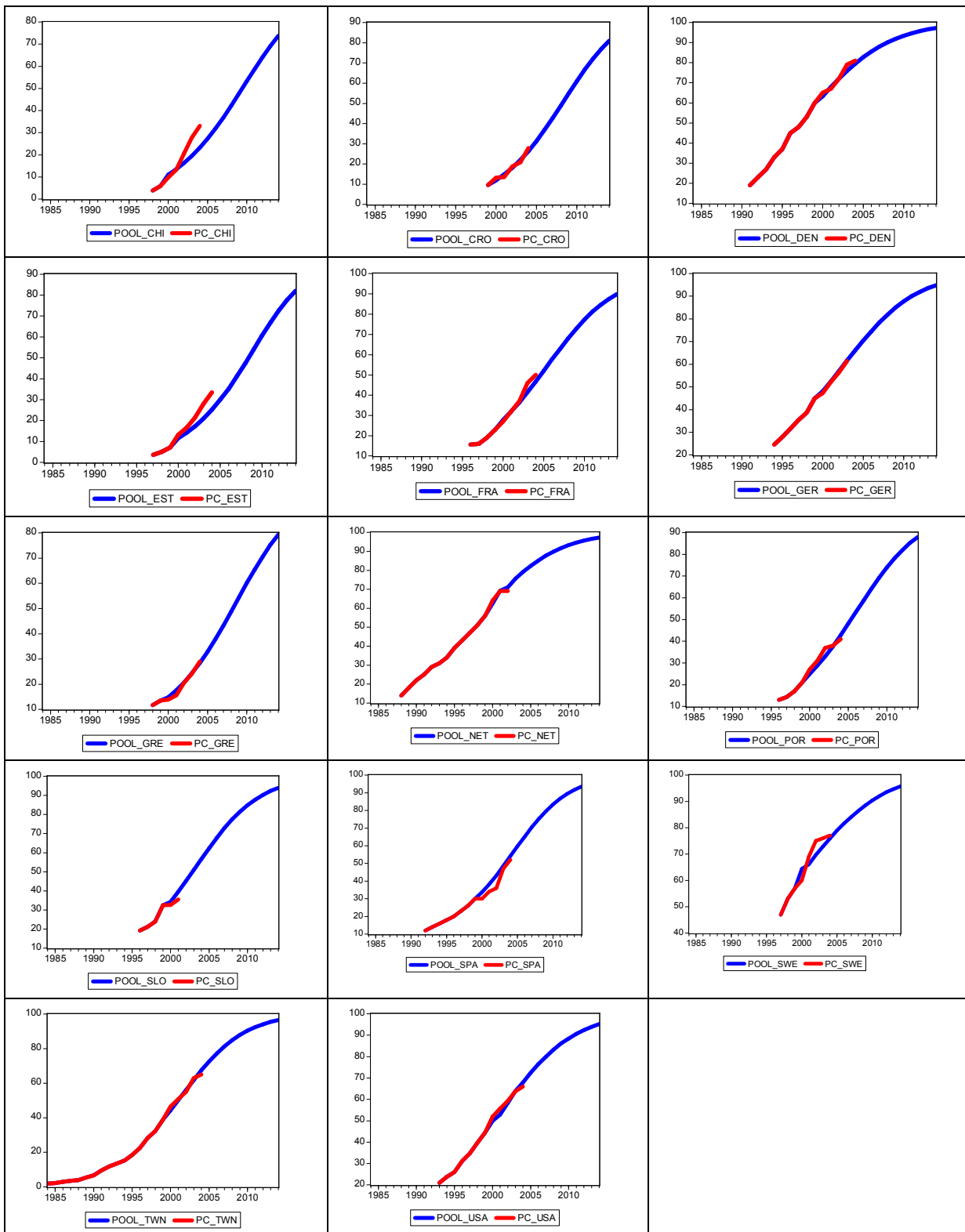
$$PC_{it} = 100 * \frac{\exp(\hat{\alpha} + \hat{\beta}_1 * TIME + \hat{\beta}_2 * UNI)}{1 + \exp(\hat{\alpha} + \hat{\beta}_1 * TIME + \hat{\beta}_2 * UNI)}$$

Aproksimacinė kreivė ir prognozė vertinant MKM su svoriais.



Grafikas 15: Logistinio modelio aproksimacinė kreivė bei modelio prognozė, MKM su svoriais.

Kaip ir kvadratinio trendo atveju, pamėginsime nubrėžti kiekvienos šalies tikrąją PC kreivę ir prognozes.



Grafikas 16: Logistinio modelio prognozės kiekvienai šaliai.

Iš grafikų matome, kad modeliavimas logistine funkcija yra iš tiesų geras. Tik dviems šalims: Kinijai ir Estijai prognozės atrodo nepakankamai gerai įvertintos.

4.3. Gompertz modelis

Gompertz modelis, pavadintas savamokslio matematiko Benjamin Gompertz vardu. Modelis skirtas procesams, kurių stebimas augimas periodo pradžioje ir pabaigoje yra lėčiausi, o tai ir yra S formos kreivė.

Aptarsime du Gompertz modelius. Pirmajame modelyje galime patys pasirinkti skverbties koeficiento augimo lubas. Kadangi iš tiesų nėra visai aišku, kaip buvo surinkti mūsų turimi duomenys, tai nagrinėsime tris variantus, t.y. lubos lygios 100%, 120% ir 140%. Nors, jei apklausa vykdoma namų ūkiams, yra tikėtina, kad skverbties koeficientas neturėtų viršyti 100%. Antrajame modelyje lubos nebus ribojamos. Palyginsime rezultatus.

4.3.1. Modelis su nustatomomis lubomis

Literatūroje dažniausiai sutinkama Gompertz modelio forma:

$$y(t) = a \exp(b * \exp(ct)), \text{ kur } a \text{ yra lubos, } c - \text{augimo greitis, } b \text{ ir } c \text{ neigiami skaičiai.}$$

Pamėginkime ištiesinti modelį. Tarkime, kad galioja prielaida, jog skverbties koeficientas negali viršyti 100%, tuomet pasižymim, kad prisisotinimo lygis $a = 100$.

$$-\log y(t) = -\log(100) - b * \exp(ct)$$

$$\log(-\log(y(t) + \log(100))) = \log(-b) + ct$$

Taigi vertinsime modelį:

$$\log(-\log(PC_{it}) + \log(100)) = \alpha + \beta_1 * TIME + \varepsilon_{it}$$

Endogeninis kintamasis: **LOG(-LOG(PC?)+LOG(100))**

Metodas: **Fixed effects: Pooled LS / Pooled EGLS (Cross-section weights). Random effects.**

Data: **2007.05.01**

Periodas: **1984 2004**

Įtraukta laiko stebėjimų: **21**

Šalių skaičius: **14**

Nesubalansuotų panelinių duomenų skaičius: **145**

Kintamieji	Koeficientai	Stand. Paklaida	t-reikšmė	p-reikšmė
C (Pooled LS)	243.1221	4.923816	49.37676	0.0000
C (Weighted)	239.9665	3.553826	67.52342	0.0000
TIME (Pooled LS)	-0.121614	0.002464	-49.35192	0.0000
TIME (Weighted)	-0.120035	0.001779	-67.48897	0.0000
C (Random)	242.6222	4.922576	49.28764	0.0000
TIME (Random)	-0.121301	0.002462	-49.27148	0.0000
Fiksuoti efektai	Pooled Least Squares	Pooled Least Squares Weighted	Atsitiktiniai efektai	
R ²	0.972028	0.983831 (0.983682)	0.941617 (0.220327)	
R ² (adjusted)	0.969016	0.982090	0.941209	
RSS	1.563035	1.553612	1.810089	

		(1.567974)	(43.56746)
Durbin-Watson	0.383683	0.716323 (0.383299)	0.331425 (0.013770)
AIC	-1.485331		

Lentelė 19: Gompertz modelis su 100% lubomis.

Hausman testas, tikrinantis kurį modelį geriau pasirinkti: fiksuotų ar atsitiktinių efektų:

Nulinė hipotezė	Atsitiktinių efektų modelis	
Chi Sq. Statistika	Chi Sq. Laisvės laipsniai	p- reikšmė
8.535009	1	0.0035

Matome, kad nulinė hipotezė yra atmetama – reikia rinktis fiksuotų efektų modelį. Taip pat pateikiama lentelė, rodanti abiejų metodų koeficientų įverčius, skirtumo variaciją ir p-reikšmes, kai nulinė hipotezė yra, kad skirtumas tarp šių įverčių yra statistiškai reikšmingas.

Nulinė hipotezė	TIME (Fixed) = TIME (Random)			
Kintamasis	Fixed	Random	Var (Diff)	P- reikšmė
TIME	-0.121614	-0.121301	0.000000	0.0035

Skirtumas yra reikšmingas. Fiksuoti efektai nėra pertekliniai:

Fiksuotų efektų testas	Statistika	Laisvės laipsniai	p- reikšmė
Cross–section F	225.977301	(13, 130)	0.0000
Cross–section Chi-sq	458.366826	13	0.0000

Pažiūrėkime kovariacinę liekanų matricą:

	CHI	CRO	DEN	EST	...	SPA	SWE	TWN	USA
CHI	0,018	-0,001	0,008	0,011	...	0,000	0,011	0,004	0,002
CRO	-0,001	0,002	0,000	-0,001	...	0,000	-0,002	0,000	0,000
DEN	0,008	0,000	0,026	0,008	...	-0,013	0,007	0,008	0,003
EST	0,011	-0,001	0,008	0,011	...	-0,002	0,010	0,004	0,003
...
SPA	0,000	0,000	-0,013	-0,002	...	0,019	-0,004	-0,004	-0,002
SWE	0,011	-0,002	0,007	0,010	...	-0,004	0,016	0,003	0,003
TWN	0,004	0,000	0,008	0,004	...	-0,004	0,003	0,018	0,002
USA	0,002	0,000	0,003	0,003	...	-0,002	0,003	0,002	0,002

Lentelė 20: Kovariacinė liekanų matrica, vertinant Gompertz modelį su 100% lubomis, MKM be svorių.

Matome, kad liekanų dispersija yra pakankamai panaši, taigi šį kartą rinksimės MKM be svorių. Pažiūrėkime liekanų koreliacijų lentelę, matome, kad liekanos yra labiau koreliuotos nei logistiniame modelyje.

	CHI	CRO	DEN	EST	...	SPA	SWE	TWN	USA
CHI	1.000	-0.201	0.393	0.810	...	-0.019	0.630	0.206	0.354
CRO	-0.201	1.000	-0.049	-0.163	...	0.021	-0.433	-0.063	-0.143
DEN	0.393	-0.049	1.000	0.457	...	-0.600	0.322	0.387	0.476
EST	0.810	-0.163	0.457	1.000	...	-0.127	0.733	0.259	0.592
...

SPA	-0.019	0.021	-0.600	-0.127	...	1.000	-0.202	-0.230	-0.321
SWE	0.630	-0.433	0.322	0.733	...	-0.202	1.000	0.193	0.465
TWN	0.206	-0.063	0.387	0.259	...	-0.230	0.193	1.000	0.350
USA	0.354	-0.143	0.476	0.592	...	-0.321	0.465	0.350	1.000

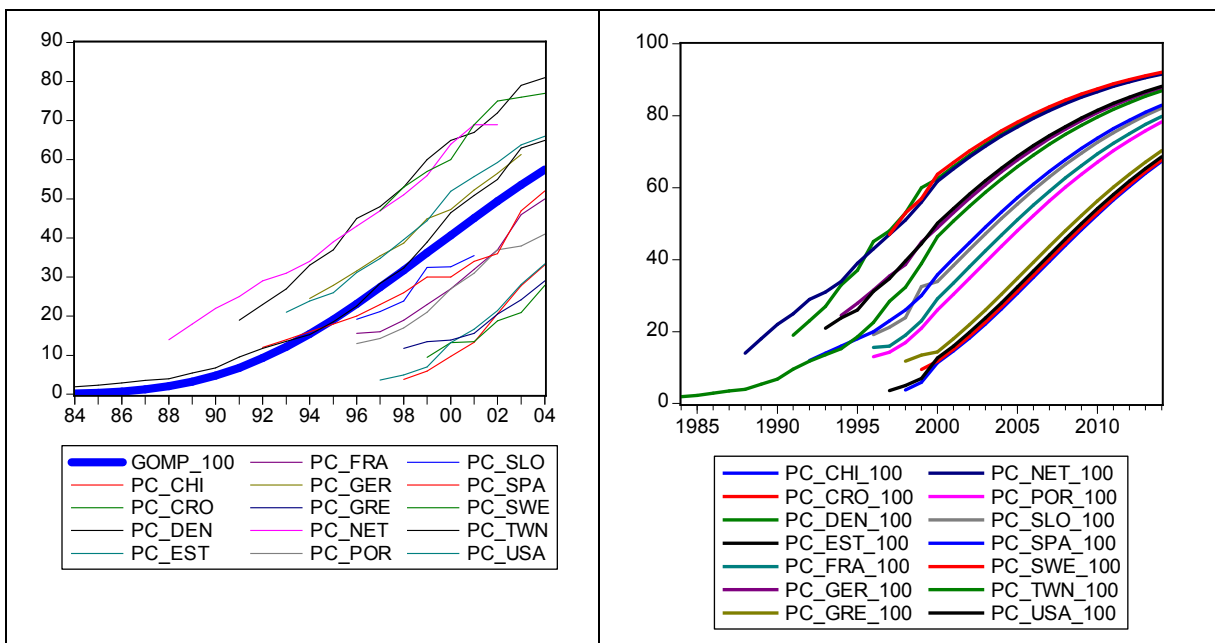
Lentelė 21: Koreliacinė liekanų matrica, vertinant Gompertz modelį su 100% lubomis, MKM be svorių.

Pamėginkime nubrėžti aproksimacinį grafiką.

$$\log(-\log(PC_t) + \log(100)) = \hat{\alpha} + \hat{\beta}_1 * TIME$$

$$\log(PC_t) - \log(100) = -\exp(\hat{\alpha} + \hat{\beta}_1 * TIME)$$

$$PC_t = 100 * \exp(-\exp(\hat{\alpha} + \hat{\beta}_1 * TIME))$$



Grafikas 17: Gompertz modelio su 100% lubomis trendo ir prognozių grafikai.

Pamėginkime pakelti lubas iki 120% ir pažiūrėti, kokie bus rezultatai. Taigi nagrinėjame modelį:

$$\log(-\log(PC_{it}) + \log(120)) = \alpha + \beta_1 * TIME + \varepsilon_{it}$$

Endogeninis kintamasis: **LOG(-LOG(PC?) + LOG(120))**

Metodas: **Fixed effects: Pooled LS / Pooled EGLS (Cross-section weights). Random effects.**

Data: **2007.05.01**

Periodas: **1984 2004**

Ištraukta laiko stebėjimų: **21**

Šalių skaičius: **14**

Nesubalansuotų panelinių duomenų skaičius: **145**

Kintamieji	Koeficientai	Stand. Paklaida	t-reikšmė	p-reikšmė
C (Pooled LS)	204.1715	3.526946	57.88902	0.0000
C (Weighted)	201.5064	2.553332	78.91902	0.0000
TIME (Pooled LS)	-0.102033	0.001765	-57.80485	0.0000

TIME (Weighted)	-0.100699	0.001278	-78.80272	0.0000
C (Random)	203.8826	3.527264	57.80192	0.0000
TIME (Random)	-0.101836	0.001764	-57.73421	0.0000
Fiksuoti efektai	Pooled Least Squares	Pooled Least Squares Weighted	Atsitiktiniai efektai	
R ²	0.979200	0.987979 (0.987825)	0.956764 (0.229247)	
R ² (adjusted)	0.976960	0.986684	0.956462	
RSS	0.801979	0.795314 (0.805501)	0.929582 (29.71797)	
Durbin-Watson	0.466361	0.837336 (0.465220)	0.402424 (0.012588)	
AIC	-2.152633			

Lentelė 22: Gompertz modelis su lubomis 120%.

Hausman testas, tikrinantis kurį modelį geriau pasirinkti: fiksuotų ar atsitiktinių efektų:

Nulinė hipotezė	Atsitiktinių efektų modelis	
Chi Sq. Statistika	Chi Sq. Laisvės laipsniai	p- reikšmė
8.783461	1	0.0030

Matome, kad nulinė hipotezė yra atmetama – reikia rinktis fiksuotų efektų modelį. Taip pat pateikiama lentelė, rodanti abiejų metodų koeficientų įverčius, skirtumo variaciją ir p-reikšmes, kai nulinė hipotezė yra, kad skirtumas tarp šių įverčių yra statistiškai reikšmingas.

Nulinė hipotezė	TIME (Fixed) = TIME (Random)			
Kintamasis	Fixed	Random	Var (Diff)	P- reikšmė
TIME	-0.102033	-0.101836	0.000000	0.0030

Skirtumas yra reikšmingas. Fiksuoti efektai nėra pertekliniai:

Fiksuotų efektų testas	Statistika	Laisvės laipsniai	p- reikšmė
Cross-section F	302.710491	(13, 130)	0.0000
Cross-section Chi-sq	499.190445	13	0.0000

Pažiūrėkime liekanų kovariacijų lentelę:

	_CHI	_CRO	_DEN	_EST	...	_SPA	_SWE	_TWN	_USA
CHI	0,019	0,000	0,003	0,013	...	-0,001	0,003	0,002	0,000
CRO	0,000	0,001	0,000	0,000	...	0,000	-0,001	0,000	0,000
DEN	0,003	0,000	0,006	0,003	...	-0,006	0,000	0,003	0,000
EST	0,013	0,000	0,003	0,013	...	-0,002	0,003	0,003	0,001
...
SPA	-0,001	0,000	-0,006	-0,002	...	0,011	-0,002	-0,003	-0,001
SWE	0,003	-0,001	0,000	0,003	...	-0,002	0,004	0,001	0,001
TWN	0,002	0,000	0,003	0,003	...	-0,003	0,001	0,008	0,001
_USA	0,000	0,000	0,000	0,001	...	-0,001	0,001	0,001	0,001

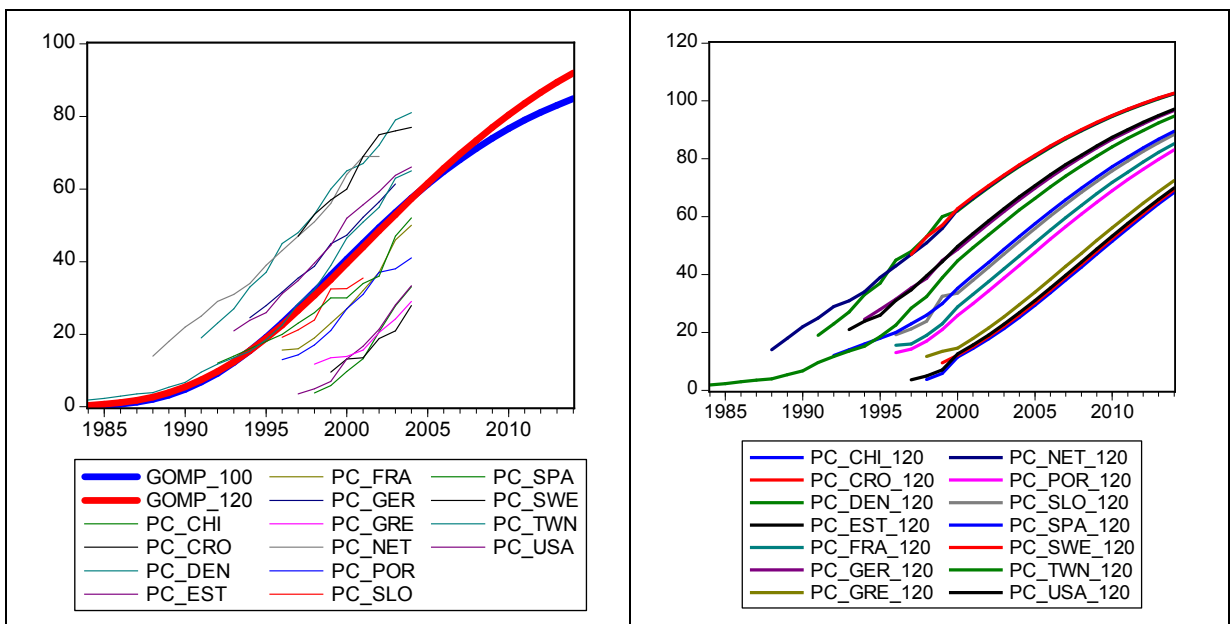
Lentelė 23: Kovariacinė liekanų matrica, vertinant Gompertz modelį su 120% lubomis, MKM be svorių.

Liekanų dispersijos mažos, pažiūrėkime liekanų koreliacijų lentelę:

	CHI	CRO	DEN	EST	...	SPA	SWE	TWN	USA
CHI	1,000	0,073	0,251	0,798	...	-0,035	0,360	0,195	0,094
CRO	0,073	1,000	0,038	0,052	...	0,024	-0,427	-0,025	-0,130
DEN	0,251	0,038	1,000	0,322	...	-0,657	0,086	0,410	0,192
EST	0,798	0,052	0,322	1,000	...	-0,149	0,470	0,262	0,358
...
SPA	-0,035	0,024	-0,657	-0,149	...	1,000	-0,252	-0,331	-0,302
SWE	0,360	-0,427	0,086	0,470	...	-0,252	1,000	0,125	0,282
TWN	0,195	-0,025	0,410	0,262	...	-0,331	0,125	1,000	0,279
USA	0,094	-0,130	0,192	0,358	...	-0,302	0,282	0,279	1,000

Lentelė 24: Koreliacinė liekanų matrica, vertinant Gompertz modelį su 120% lubomis, MKM be svorių.

Taigi renkamės MKM su svoriais, pažiūrėkime grafikus. Aproximaciniame grafike padidinau intervalą, kad matytųsi, jog lubų keitimas iš tiesų turi įtakos:



Grafikas 18: Gompertz modelio su 120% lubomis trendo ir prognozių grafikai.

Pamėginkime dar daugiau pakelti lubas iki 140% ir pažiūrėti, kokie bus rezultatai.

$$\log(-\log(PC_{it}) + \log(140)) = \alpha + \beta_1 * TIME + \varepsilon_{it}$$

Endogeninis kintamasis: **LOG(-LOG(PC?) + LOG(140))**

Metodas: **Fixed effects: Pooled LS / Pooled EGLS (Cross-section weights). Random effects.**

Data: **2007.05.02**

Periodas: **1984 2004**

Įtraukta laiko stebėjimų: **21**

Šalių skaičius: **14**

Nesubalansuotų panelinių duomenų skaičius: **145**

Kintamieji	Koeficientai	Stand. Paklaida	t-reikšmė	p-reikšmė
C (Pooled LS)	181.3782	3.033220	59.79723	0.0000
C (Weighted)	178.6006	2.153409	82.93853	0.0000
TIME (Pooled LS)	-0.090565	0.001518	-59.65935	0.0000

TIME (Weighted)	-0.089175	0.001078	-82.74429	0.0000
C (Random)	181.1388	3.033665	59.70955	0.0000
TIME (Random)	-0.090399	0.001517	-59.59019	0.0000
Fiksuoti efektai	Pooled Least Squares	Pooled Least Squares Weighted	Atsitiktiniai efektai	
R ²	0.980334	0.989261 (0.989053)	0.959314 (0.234941)	
R ² (adjusted)	0.978216	0.988104	0.959029	
RSS	0.593162	0.585666 (0.596988)	0.687610 (0.012572)	
Durbin-Watson	0.489016	0.875863 (0.486974)	0.421917 (0.012588)	
AIC	-2.454249			

Lentelė 25: Gompertz modelis su lubomis 140%.

Hausman testas, tikrinantis kurį modelį geriau pasirinkti: fiksuotų ar atsitiktinių efektų:

Nulinė hipotezė	Atsitiktinių efektų modelis	
Chi Sq. Statistika	Chi Sq. Laisvės laipsniai	p- reikšmė
8.870585	1	0.0029

Matome, kad nulinė hipotezė yra atmetama – reikia rinktis fiksuotų efektų modelį. Taip pat pateikiama lentelė, rodanti abiejų metodų koeficientų įverčius, skirtumo variaciją ir p-reikšmes, kai nulinė hipotezė yra, kad skirtumas tarp šių įverčių yra statistiškai reikšmingas.

Nulinė hipotezė	TIME (Fixed) = TIME (Random)			
Kintamasis	Fixed	Random	Var (Diff)	P- reikšmė
TIME	-0.090565	-0.090399	0.000000	0.0029

Skirtumas iš tiesų yra reikšmingas. Fiksuoti efektai nėra pertekliniai:

Fiksuotų efektų testas	Statistika	Laisvės laipsniai	p- reikšmė
Cross-section F	318.272569	(13, 130)	0.0000
Cross-section Chi-sq	506.232580	13	0.0000

Pažiūrėkime liekanų kovariacijų lentelę:

	_CHI	_CRO	_DEN	_EST	...	_SPA	_SWE	_TWN	_USA
_CHI	0.019	0.001	0.001	0.013	...	-0.001	0.000	0.002	0.000
_CRO	0.001	0.001	0.000	0.001	...	0.000	-0.001	0.000	0.000
_DEN	0.001	0.000	0.002	0.001	...	-0.003	0.000	0.001	0.000
_EST	0.013	0.001	0.001	0.013	...	-0.002	0.000	0.002	0.000
...
_SPA	-0.001	0.000	-0.003	-0.002	...	0.008	-0.001	-0.002	0.000
_SWE	0.000	-0.001	0.000	0.000	...	-0.001	0.002	0.000	0.000
_TWN	0.002	0.000	0.001	0.002	...	-0.002	0.000	0.005	0.000
_USA	0.000	0.000	0.000	0.000	...	0.000	0.000	0.000	0.001

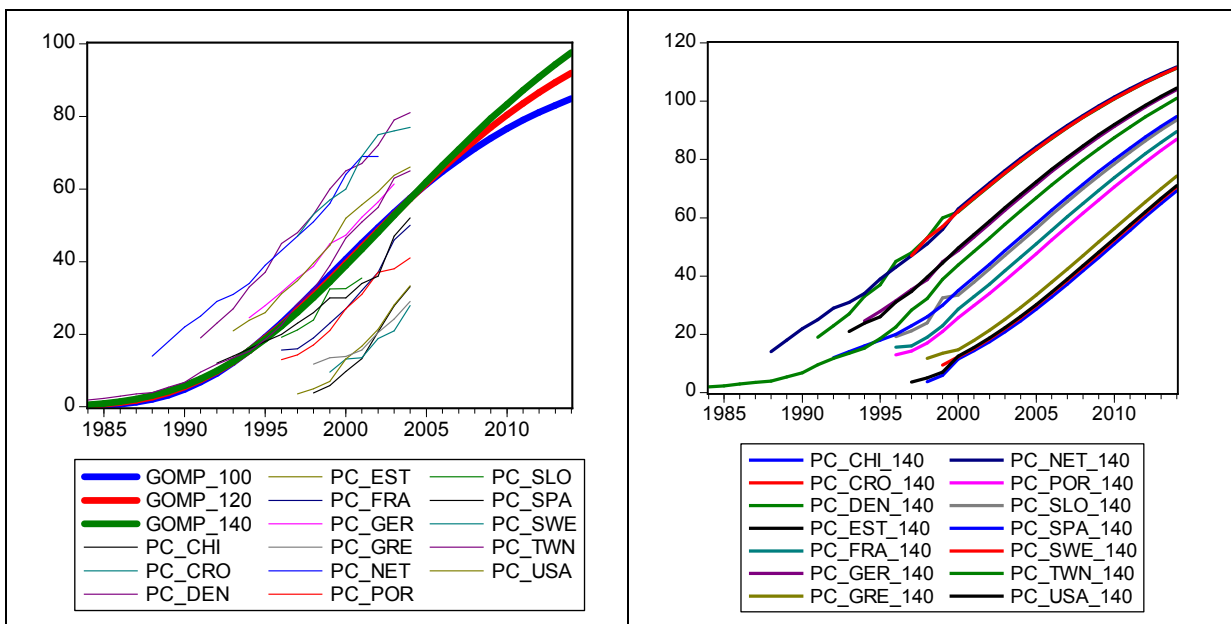
Lentelė 26: Kovariacinė liekanų matrica, vertinant Gompertz modelį su lubomis 140%, MKM be svorių.

Matome, kad dispersija nėra labai skirtinga. Pažiūrėkime liekanų koreliacijų lentelę:

	CHI	CRO	DEN	EST	...	SPA	SWE	TWN	USA
CHI	1,000	0,215	0,094	0,791	...	-0,053	0,022	0,180	-0,131
CRO	0,215	1,000	0,024	0,160	...	0,018	-0,499	-0,012	-0,196
DEN	0,094	0,024	1,000	0,157	...	-0,634	-0,070	0,374	0,015
EST	0,791	0,160	0,157	1,000	...	-0,169	0,097	0,253	0,111
...
SPA	-0,053	0,018	-0,634	-0,169	...	1,000	-0,247	-0,402	-0,194
SWE	0,022	-0,499	-0,070	0,097	...	-0,247	1,000	0,036	0,232
TWN	0,180	-0,012	0,374	0,253	...	-0,402	0,036	1,000	0,153
USA	-0,131	-0,196	0,015	0,111	...	-0,194	0,232	0,153	1,000

Lentelė 27: Koreliacinė liekanų matrica, vertinant Gompertz modelį su lubomis 140%, MKM be svorių.

Taigi renkamės MKM be svorių, pažiūrėkime grafikus, jie atrodo labai panašiai kaip ir prieš tai nagrinėtų modelių.



Grafikas 19: Gompertz modelio su 140% lubomis trendo ir prognozių grafikai.

4.3.2. Modelis, kai lubos nėra nustatomos iš anksto

Netiesinė modelio forma:

$$Y_t = m * \exp(-\exp(-(a + b * t))) + \varepsilon_t,$$

kur Y_t yra kumuliacinis pirkėjų ar vartotojų skaičius momentu t , m yra prisisotinimo lygis, a ir b yra nekintantys koeficientai. Pamėginsime išvesti tiesinę modelio formą:

$$\ln Y_t = \ln m - \exp(-(a + b * t)),$$

$$\ln Y_{t-1} = \ln m - \exp(-(a + b * (t-1))) = \ln m - \exp(b) * \exp(-(a + bt)),$$

Atimame lygtis vieną iš kitos:

$$\ln Y_t - \ln Y_{t-1} = -\exp(-(a + b * t))(1 - \exp(b)),$$

Vietoje $-\exp(-(a + b * t))$ įsistatome $(\ln Y_{t-1} - \ln m) / \exp(b)$ ir gauname:

$$\ln Y_t - \ln Y_{t-1} = (1 - \exp(b)) * \frac{(\ln Y_{t-1} - \ln m)}{\exp(b)},$$

Pažymėję $\frac{(1 - \exp(b))}{\exp(b)} = \varphi$, turime tiesinę Gompertz modelio funkciją:

$$\ln(Y_{it}) - \ln(Y_{it-1}) = \varphi(\ln m - \ln Y_{it-1}) = \varphi \ln m - \varphi \ln Y_{it-1}$$

čia φ yra augimo parametras. Taigi tarsime, kad φ yra toks pats visoms šalims, o prisisotinimo lygis m skiriasi tarp šalių. Šiuo atveju lubos nėra lygios nei 100, nei 120, nei 140. Vertiname modelį:

$$\ln(PC_{it}) - \ln(PC_{it-1}) = \alpha + \beta_1 \ln PC_{it-1} + \delta_i + \varepsilon_{it}$$

Endogeninis kintamasis: **LOG(PC?) - LOG(PC?(-1))**

Metodas: **Fixed effects: Pooled LS / Pooled EGLS (Cross-section weights). Random effects.**

Data: **2007.05.10**

Periodas: **1984 2004**

Įtraukta laiko stebėjimų: **20**

Šalių skaičius: **14**

Nesubalansuotų panelinių duomenų skaičius: **131**

Kintamieji	Koeficientai	Stand. Paklaida	t-reikšmė	p-reikšmė
C (Pooled LS)	0.336151	0.034464	0.034464	0.0000
C (Weighted)	0.357707	0.029204	12.24873	0.0000
LOG(PC?(-1)) (Pooled LS)	-0.057749	0.010623	-5.436212	0.0000
LOG(PC?(-1)) (Weighted)	-0.064502	0.009056	-7.122845	0.0000

C (Random)	0.380113	0.032884	11.55909	0.0000
LOG(PC?(-1)) (Random)	-0.069793	0.009628	-7.248990	0.0000
Fiksuoti efektai	Pooled Least Squares	Pooled Least Squares Weighted	Atsitiktiniai efektai	
R ²	0.577688	0.564054 (0.560438)	0.319791 (0.372123)	
R ² (adjusted)	0.526719	0.511440	0.314518	
RSS	0.574967	0.572225 (0.576970)	0.672501 (0.854838)	
Durbin-Watson	2.351115	2.177184 (2.326900)	1.986003 (1.562388)	
AIC	-2.361754			

Lentelė 28: Gompertz modelis be lubų.

Nors Durbin-Watson statistika yra arti 2, tačiau yra žinoma, kad šis testas nėra tinkamas, kai tarp paaiškinančiųjų kintamųjų yra endogeninių kintamųjų su pavėlinimu. Hausman testas, tikrinantis kurį modelį geriau pasirinkti: fiksuotų ar atsitiktinių efektų:

Nulinė hipotezė	Atsitiktinių efektų modelis	
Chi Sq. Statistika	Chi Sq. Laisvės laipsniai	p- reikšmė
7.196564	1	0.0073

Matome, kad nulinė hipotezė yra atmetama – reikia rinktis fiksuotų efektų modelį. Taip pat pateikiama lentelė, rodanti abiejų metodų koeficientų įverčius, skirtumo variaciją ir p-reikšmes, kai nulinė hipotezė yra, kad skirtumas tarp šių įverčių yra statistiškai reikšmingas.

Nulinė hipotezė	TIME (Fixed) = TIME (Random)			
Kintamasis	Fixed	Random	Var (Diff)	P- reikšmė
TIME	-0.057749	-0.069793	0.000020	0.0073

Skirtumas yra reikšmingas. Fiksuoti efektai nėra pertekliniai:

Fiksuotų efektų testas	Statistika	Laisvės laipsniai	p- reikšmė
Cross-section F	4.127862	(13, 116)	0.0000
Cross-section Chi-sq	49.808722	13	0.0000

Kovariacinė liekanų matrica:

	_CHI	_CRO	_DEN	_EST	...	_SPA	_SWE	_TWN	_USA
CHI	0.007	0.003	0.001	0.006	...	-0.002	0.000	0.001	0.001
CRO	0.003	0.016	0.000	0.004	...	-0.003	-0.002	0.000	0.000
DEN	0.001	0.000	0.002	0.000	...	0.000	0.000	0.000	0.001
EST	0.006	0.004	0.000	0.015	...	-0.003	-0.001	0.001	0.002
...
SPA	-0.002	-0.003	0.000	-0.003	...	0.004	0.000	0.000	-0.001
SWE	0.000	-0.002	0.000	-0.001	...	0.000	0.002	0.000	0.000
TWN	0.001	0.000	0.000	0.001	...	0.000	0.000	0.004	0.000
_USA	0.001	0.000	0.001	0.002	...	-0.001	0.000	0.000	0.001

Lentelė 29: Kovariacinė liekanų matrica, vertinant Gompertz modelį be lubų, MKM be svorių.

Matome, kad liekanų dispersija nėra didelė. Koreliacinė liekanų matrica:

	CHI	CRO	DEN	EST	...	SPA	SWE	TWN	USA
CHI	1.000	0.321	0.188	0.586	...	-0.363	0.120	0.119	0.380
CRO	0.321	1.000	0.046	0.255	...	-0.348	-0.387	-0.019	0.077
DEN	0.188	0.046	1.000	0.087	...	-0.038	-0.120	0.004	0.419
EST	0.586	0.255	0.087	1.000	...	-0.379	-0.294	0.147	0.501
...
SPA	-0.363	-0.348	-0.038	-0.379	...	1.000	-0.204	0.017	-0.326
SWE	0.120	-0.387	-0.120	-0.294	...	-0.204	1.000	-0.052	-0.015
TWN	0.119	-0.019	0.004	0.147	...	0.017	-0.052	1.000	0.188
USA	0.380	0.077	0.419	0.501	...	-0.326	-0.015	0.188	1.000

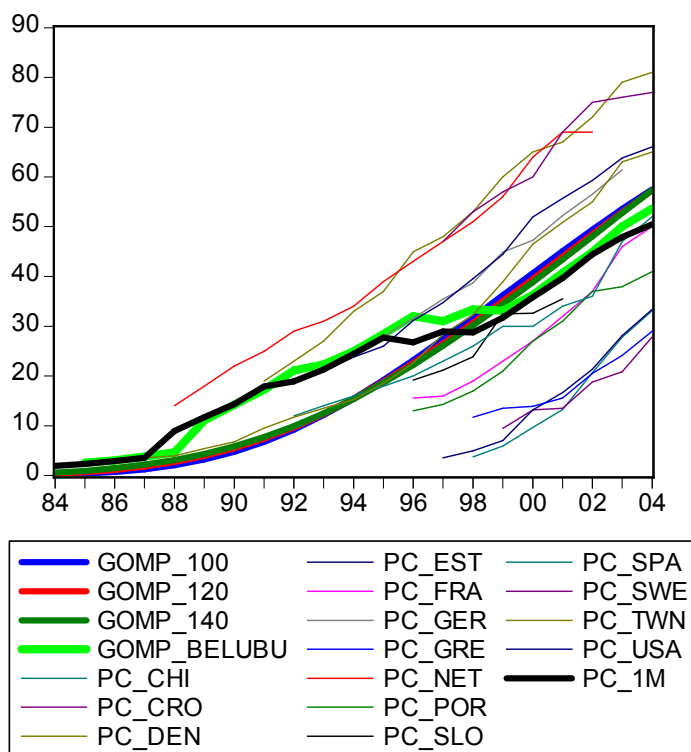
Lentelė 30: Koreliacinė liekanų matrica, vertinant Gompertz modelį be lubų, MKM be svorių.

Pasirinkime MKM be svorių. Paskaičiavę įvertinto modelio vidurkinės reikšmes pažiūrėkime kaip atrodo grafikas su modelio įvertinta kreive:

$$\ln(PC_t) - \ln(PC_{t-1}) = \ln\left(\frac{PC_t}{PC_{t-1}}\right) = \hat{\alpha} - \hat{\beta}_1 \ln PC_{t-1},$$

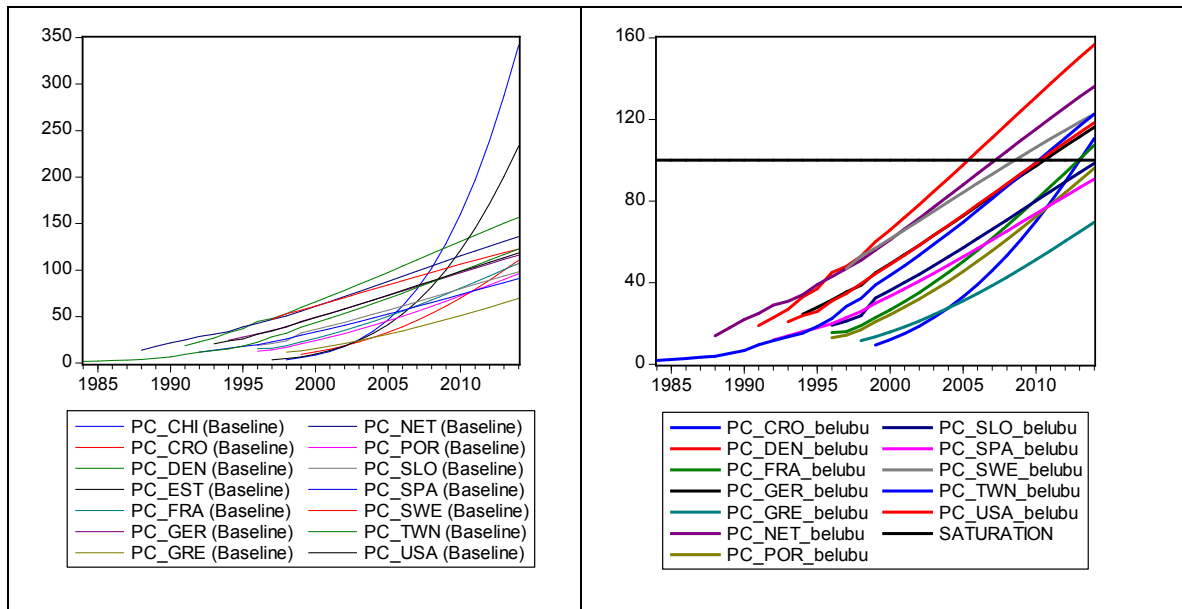
$$\frac{PC_t}{PC_{t-1}} = \exp(\hat{\alpha} - \hat{\beta}_1 \ln PC_{t-1}),$$

$$PC_t = PC_{t-1} * \exp(\hat{\alpha} - \hat{\beta}_1 \ln PC_{t-1})$$



Grafikas 20: Gompertz modelio tendas.

Pabraižykime prognozes. Pirmu atveju išsiskiriantys stebėjimai yra Kinija ir Estija, kaip ir logistiniame modelyje, šios šalys elgiasi skirtingai nei visos kitos. Antrame grafike nubrėžtas grafikas be Kinijos ir Estijos, matome, kad, didžiausias prognozuojamas augimas yra Danijoje ir Olandijoje, mažiausias – Graikijoje. Šio modelio prognozės akivaizdžiai viršija 100% .

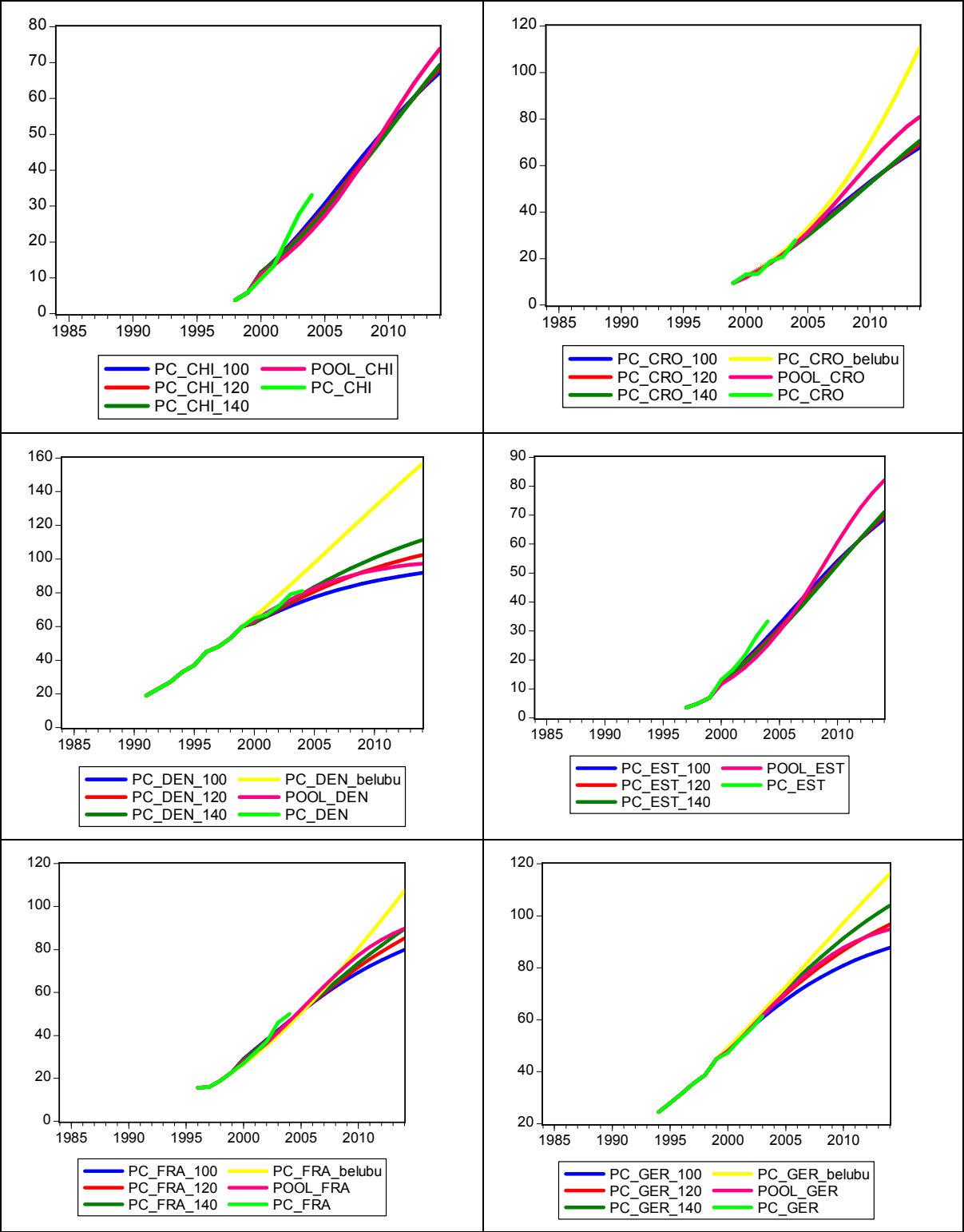


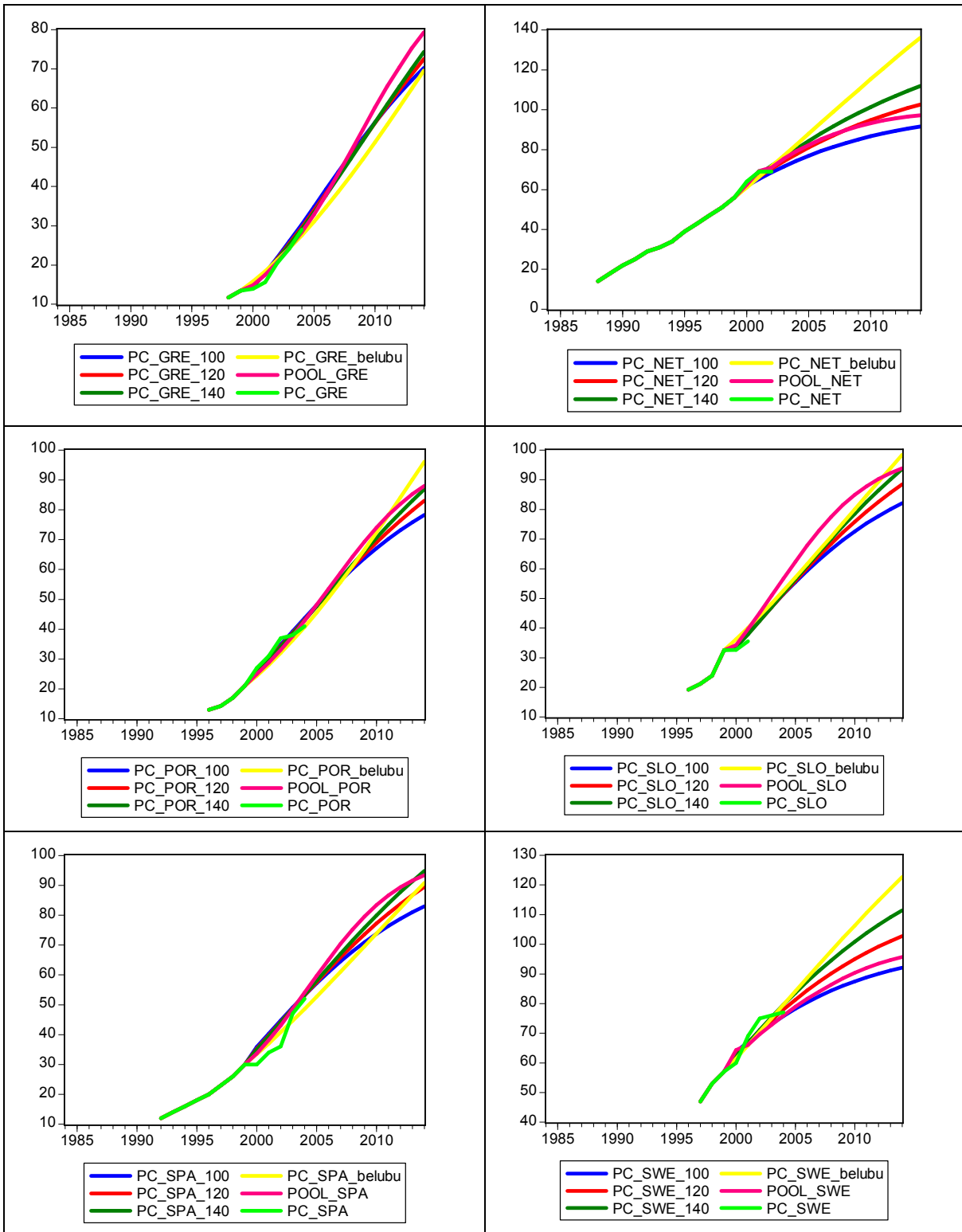
Grafikas 21: Gompertz modelio be lubų prognozės.

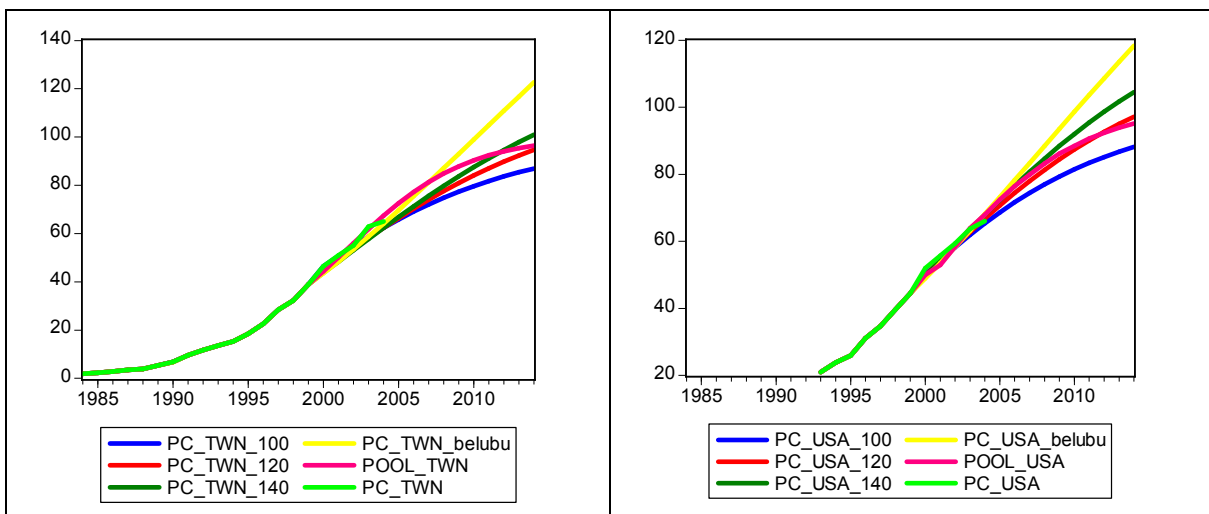
Kaip jau minėta anksčiau, šis modelis, ne taip, kaip logistinis modelis, leidžia skverbties koeficientui peržengti 100% ribą.

Vėl pamėginsime nubrėžti prognozes atskiroms šalims, Gompertz modelio be lubų prognozės nevaizduosime Kinijai ir Estijai. Kadangi šiuo atveju brėžiame daugiau kreivių, padidinsime grafikus ir kartu nubrėšime ir logistinio modelio kreives. Taigi:

- PC_CRO_100 – mėlyna kreivė yra Gompertz modelis su lubomis 100%
- PC_CRO_120 – raudona kreivė yra Gompertz modelis su lubomis 120%
- PC_CRO_140 – tamsiai žalia kreivė yra Gompertz modelis su lubomis 140%
- PC_CRO_belubu – geltona kreivė yra Gompertz modelis be lubų.
- POOL_CRO – rožinė kreivė yra logistinis modelis su lubomis 100%
- PC_CRO – šviesiai žalia kreivė yra PC duomenys.







Grafikas 22: Logistinio ir Gompertz modelio su 100%, 120% ir 140% lubomis prognozės kiekvienai šaliai.

Modelis be nustatytų lubų akivaizdžiai pervertina skverbties koeficiento augimą. Tiesą sakant lubų parinkimas yra sudėtingas dalykas, iš tiesų tai priklauso nuo pradinės prielaidos. Paskaičiuokite kiekvienai šaliai šaknį iš kvadratinių paklaidų vidurkio (Root Mean Square Error – RMSE) ir vidutinę absoliutinę paklaidą (Mean Absolute Error – MAE). Matome, kad didžiausios paklaidos yra Kinijai ir Estijai. Kinijai per 7 metus skverbties koeficientas padidėjo nuo 3.78 iki 33.11, tai reiškia, kad kiekvienais metais jis augo po maždaug 40%. Estijai panašiai, per 8 metus koeficientas padidėjo nuo 3.60 iki 33.40, tai reiškia, kad kiekvienais metais jis augo po 30%. Tuo tarpu kitoms šalims augimas buvo kuklesnis, vidutiniškai 10 - 20%. Taigi Kinijoje ir Estijoje buvo ypatingai didelis kompiuterių skverbties augimas. Daugeliu atveju mažiausios paklaidos yra Gompertz modelio su 140% lubomis.

Šalis	RMSE				MAE			
	Logist	100	120	140	Logist	100	120	140
CHI	0.409	0.133	0.138	0.137	0.355	0.123	0.126	0.124
CRO	0.083	0.042	0.035	0.034	0.080	0.037	0.030	0.029
DEN	0.082	0.161	0.079	0.050	0.067	0.134	0.067	0.042
EST	0.373	0.104	0.114	0.116	0.347	0.092	0.102	0.104
FRA	0.100	0.064	0.054	0.048	0.079	0.059	0.049	0.042
GER	0.023	0.030	0.025	0.025	0.019	0.026	0.022	0.023
GRE	0.082	0.077	0.057	0.048	0.067	0.066	0.049	0.040
NET	0.060	0.071	0.048	0.041	0.051	0.065	0.046	0.038
POR	0.101	0.045	0.036	0.033	0.089	0.038	0.034	0.031
SLO	0.114	0.057	0.046	0.040	0.099	0.050	0.040	0.035
SPA	0.145	0.140	0.106	0.089	0.122	0.113	0.085	0.072
SWE	0.166	0.126	0.061	0.043	0.156	0.118	0.055	0.035
TWN	0.069	0.133	0.090	0.069	0.059	0.105	0.076	0.060
USA	0.058	0.045	0.030	0.025	0.048	0.042	0.027	0.021

Lentelė 31: RMSE ir MSE palyginimas visoms šalims logistiniam ir Gompertz modeliams su 100%, 120% ir 140% lubomis.

Taip pat pateiksiu lentelę su pagrindinėmis modelių statistikomis:

	R ²	R ² (adjusted)	RSS	Durbin-Watson
Logistinis	0.980	0.978	3.359	0.513
Gompertz_100	0.972	0.969	1.563	0.383
Gompertz_120	0.979	0.976	0.801	0.466
Gompertz_140	0.980	0.978	0.593	0.489
Gompertz_belubu	0.577	0.527	0.574	2.351*

Lentelė 32: pagrindinių statistikų palyginimas logistiniams ir Gompertz modeliams su 100%, 120% ir 140% lubomis bei be lubų. * reiškia, kad Durbin-Watson statistika paskutiniame modelyje blogai įvertinta, nes dešinėje modelio pusėje yra endogeninis kintamasis su pavėlinimu.

Taigi, jei daroma prielaida, kad skverbties koeficientas gali peržengti 100% augimo ribą, geriausias modelis būtų Gompertz su 140 nustatytomis maksimaliomis lubomis. Tačiau, jei lubos vis dėlto yra 100%, geriausiai tinkantis modelis yra logistinis.

4.4. Bass modelis

Bass difuzijos modelį išvystė Frank Bass (1926 – 2006, JAV; Jis laikomas vienu iš marketingo mokslo įkūrėjų). Modelis paaiškina procesą, kaip nauji produktai įsisavinami sąlygojant sąveikai tarp produkto naudotojų ir potencialių pirkėjų. Modelis taip pat plačiai naudojamas pardavimų prognozavimui.

Kadangi įmonės visokiais būdais stengiasi įsitvirtinti atsirandančiose rinkose, joms yra labai svarbu įvertinti santykinį rinkos plėtros patrauklumą skirtingose šalyse. Bass teigia, kad rinkos patrauklumas yra funkcija nuo galimo rinkos potencialo ir greičio, kuriuo produktas pasklinda rinkoje. Taigi geresnis rinkos potencialą sąlygojančių veiksnių ir difuzijos greičio skirtingose šalyse supratimas yra ypatingai svarbus firmoms, svarstančioms savo plėtimosi rinkoje strategijas.

Reikia atsižvelgti ir į kai kuriuos makroekonominis rodiklius, kaip pavyzdžiui, išsilavinimas, urbanizacijos lygis ar tarptautinė prekyba, kurių kilimas skatina ir skverbties potencialą. Nors verslininkai tokių rodiklių praktiškai neįtakoja, tačiau vis tiek turi į juos atsižvelgti, vertindami skirtingas tarptautines rinkas ir darydami jautrumo analizę.

Taip pat ši analizė naudinga verslininkams, kurie nori apjungti informaciją apie praeities difuzijas tarp skirtingų šalių ir skirtingų produktų, kad gauti geresnes prognozes. Bass modelio vertinimo technologijoje yra efektyviai apjungiamą informaciją tarp skirtingų produktų ir šalių naudojant hierarchinę Bayes įvertinimo metodologiją. Informacijos apie šalis ir produktus sujungimas leidžia pagerinti prognozę. Šis metodas duoda daugiausiai naudos prognozuojant pardavimus, kai produktas yra naujai paleidžiamas į kažkokią naują rinką. Praeities patyrimas tam pačiam produktui kitose šalyse (product effects) yra naudingesnis aiškinant išorinių ir vidinių įtakų koeficientus ir artėjimo prie pardavimų pikų greitį, o kitų produktų praeitis toje pačioje šalyje (country effects) geriau paaiškina skverbti.

Metodologija

Naujo produkto plitimas gali būti charakterizuotas naudojant Bass formulę. Tiesinė forma atrodo taip:

$$Y_t - Y_{t-1} = p * m + (q - p) * Y_{t-1} + \frac{q}{m} * Y_{t-1}^2 + \varepsilon_t$$

Tiesinis Bass modelis gali būti gerai interpretuojamas tik naudojant atsiktinių koeficientų modelį (paminėtą 2.2.3. skyriuje). Deja, E-Views paketas to negali realizuoti, todėl čia aptarsime tik šio metodo teorinius pagrindus. Pirmiausiai šiek tiek plačiau aprašysime netiesinę Bass modelio formą.

$$S_{pr,c}(t) - S_{pr,c}(t-1) = (p_{pr,c} + \left(\frac{q_{pr,c}}{m_{pr,c}}\right) * S_{pr,c}(t-1)) * (m_{pr,c} - S_{pr,c}(t-1))$$

Kur p_r – produktas, c – šalis, $S_{pr,c}(t)$ – sukaupti kažkokio produkto pardavimai šalyje c momentu t . Taigi $S_{pr,c}(t)$ yra skaičius įmonių arba žmonių, kurie naudoja naująjį produktą.

m – rinkos skverbties potencialas (market penetration potencial) arba lubos produktui. Yra žinoma, kad ne visi, sužinoję apie naują produktą, jį pirsks. Ekonomikos teorija teigia, kad vartotojai, kurie priima naują produktą yra tie, kurie turi:

1. Galimybę pirkti.
2. Norą mokėti.
3. Priėjimą prie produkto.

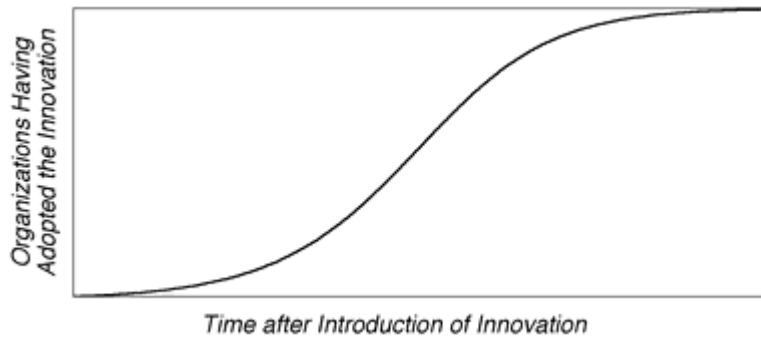
p – išorinės įtakos koeficientas, taip pat dažnai vadinamas inovacijos koeficientu. Tai tikimybė, kad žmogus, kuris dar nenaudoja produkto, pradės jį naudoti dėl visuomenės informavimo ir kitų išorinių priemonių poveikio. Šis koeficientas nepriklauso nuo:

1. Žmonių, priimančių produktą, skaičiaus šalyje. Tokios įtakos stiprėjimas priklauso nuo vartotojų priėjimo prie informacijos susijusios su preke (kuri gali būti pasiekama be tiesioginio bendravimo).
2. Vartotojų polinkio ir galimybių gauti informaciją iš netiesioginės komunikacijos šaltinių.

q – vidinės įtakos koeficientas, taip pat dažnai vadinamas imitacijos koeficientu. Tai tikimybė, kad žmogus, kuris dar nenaudoja produkto, pradės jį naudoti dėl „žodis iš burnos“ proceso ar kitų žmonių, kurie jau naudoja produktą, įtakos. Skaičius priklauso nuo žmonių, priimančių produktą, skaičiaus šalyje, nes jie įtakoja būsimus vartotojus. Taigi p ir q nurodo prekės priėmimo būdą ir greitį. Koeficientas bus didesnis, jei:

1. Populiacija bus homogeninė.
2. Vartotojų, įsigijusių produktą, skatinimas bus didesnis.

Standartinė Bass kreivė su vidutinėmis reikšmėmis $p = 0.03$ ir $q = 0.38$, atrodo taip:



Grafikas 23: Bass difuzijos modelio grafikas.

Modelio prielaidos:

1. Liekanos modeliuojamos multiplikatyviai kaip įprasta modeliuojant paklausą, taip sumažinamas heteroskedastiškumas ir apsisaugoma nuo galimų neigiamų paklausos reikšmių.
2. Leidžiama liekanų koreliuotumo galimybė, taip pat liekanų autokoreliacija.

Perrašykime mūsų Bass difuzijos modelį kitu pavidalu, įtraukdami multiplikatyvias koreliuotas paklaidas $\varepsilon_{pr,c}(t)$, gauname:

$$S_{pr,c}(t) - S_{pr,c}(t-1) = \alpha_{pr,c} (F_{pr,c}(t) - F_{pr,c}(t-1)) * \exp(\varepsilon_{pr,c}(t))$$

Norint užtikrinti, kad visi trys parametrai bus teigiami, pasirenkama eksponentinė netiesinė transformacija. Tuo pačiu transformuotos reikšmės gali būti realios, tuomet variacijos įvertinimui galima būtų naudoti normalųjį pasiskirstymą. Taigi transformuotos reikšmės su “*“:

$$\alpha_{pr,c} = \exp(\alpha_{pr,c}^*), \quad p_{pr,c} = \exp(p_{pr,c}^*) \quad \text{ir} \quad q_{pr,c} = \exp(q_{pr,c}^*).$$

Transformavę koeficientus, juos išskiriam į tris komponentes:

- Komponentę, kuri bendra kažkokiam produktui, atsižvelgiant į visas šalis.
- Komponentę, kuri bendra kažkokioms šalims, visų produktų atžvilgiu.
- Komponentę, kuri yra specifinė konkrečiai šaliai ir konkrečiam produktui.

$$\begin{bmatrix} \alpha_{pr,c}^* \\ p_{pr,c}^* \\ q_{pr,c}^* \end{bmatrix} = \begin{bmatrix} \alpha_{pr}^* + \alpha_c^* \\ p_{pr}^* + q_c^* \\ q_{pr}^* + q_c^* \end{bmatrix} + \begin{bmatrix} v_{\alpha,pr,c} \\ v_{p,pr,c} \\ v_{q,pr,c} \end{bmatrix}$$

Paskutiniąją komponentę skaidom į regresorius ir paklaidas:

$$\begin{bmatrix} v_{\alpha,pr,c} \\ v_{p,pr,c} \\ v_{q,pr,c} \end{bmatrix} = \begin{bmatrix} X'_{\alpha,pr,c} \beta_{\alpha,pr,c} \\ X'_{p,pr,c} \beta_{p,pr,c} \\ X'_{q,pr,c} \beta_{q,pr,c} \end{bmatrix} + \begin{bmatrix} \pi_{\alpha,pr,c} \\ \pi_{p,pr,c} \\ \pi_{q,pr,c} \end{bmatrix}, \quad \text{kur} \begin{bmatrix} \pi_{\alpha,pr,c} \\ \pi_{p,pr,c} \\ \pi_{q,pr,c} \end{bmatrix} \sim MVN(0, \lambda)$$

Tuomet galime išskaidyti tik pagal šalį arba tik pagal produktą, mūsų atveju turime tik vieno produkto pardavimų duomenis daugelyje šalių, taigi skaidysime pagal šalis:

$$\begin{bmatrix} \alpha_c^* \\ p_c^* \\ q_c^* \end{bmatrix} = \begin{bmatrix} X'_{\alpha,c} \beta_{\alpha,c} \\ X'_{p,c} \beta_{p,c} \\ X'_{q,c} \beta_{q,c} \end{bmatrix} + \begin{bmatrix} \pi_{\alpha,c} \\ \pi_{p,c} \\ \pi_{q,c} \end{bmatrix}, \quad \text{kur} \begin{bmatrix} \pi_{\alpha,c} \\ \pi_{p,c} \\ \pi_{q,c} \end{bmatrix} \sim MVN(0, \gamma_c)$$

Tai yra lanksti ir nesudėtinga heterogeniškumo tarp šalių specifikacija.

Įvertinimas

Visos šios lygtys vertinamos naudojant hierarchinę Bayes vertinimo metodologiją, kuri laikoma viena iš galingiausių modernios statistikos technologijų. Dauguma žingsnių gali būti įvertinti naudojant Markovo grandines (Markov Chain Monte Carlo - MCMC) metodus, tokius kaip Gibbs modelį, nes sąlyginiai pasiskirstymai turi jungtines pradines reikšmes (priors) ir gerai apibrėžtus pasiskirstymus, išskyrus parametrus bei autokoreliuotumą. Jiems įvertinti naudojami Metropolis – Hastings žingsniai. Būtų įdomu palyginti mūsų gautus rezultatus su rezultatais gautais vertinant Bass difuzijos modelį, tačiau dėl šio modelio įgyvendinimo sudėtingumo šiame darbe tai nebus padaryta. Daugiau apie šias procedūras galima paskaityti knygose: Gary Cooper. *Bayesian Econometrics*. Wiley (2003). Peter E. Rossi. Greg M. Allenby. Robert McCulloh. *Bayesian Statistics and Marketing*. Wiley (2005).

5. Išvados

Tarptautinio masto informacijos, tokios, kaip inovacijų difuzija skirtingoms šalims, naudojimas leidžia patikimiau prognozuoti skverbties koeficiento augimą atskiroms šalims. Sujungtų duomenų naudojimas, padidina turimos informacijos kiekį, kuris nėra didelis dėl trumpų laiko eilučių.

Šiame darbe mėginome keletu skirtingų būdų modeliuoti skverbties koeficientą ir surasti geriausiai tinkantį modelį kompiuterių skverbties prognozavimui. Pirmiausia mėginome aproksimuoti tiesiniu trendu, tačiau šis būdas galėtų būti laikomas naiviu dėl to, kad pardavimų augimai negali visą laiką vien tik augti, kažkuriame taške jie turėtų pasiekti rinkos prisisotinimo tašką. Tuomet bandėme aproksimuoti kvadratinu trendu, tačiau išbrėžus prognozių grafiką matėme, kad jis blogai aproksimuoja mūsų analizuojamus duomenis ir šis modelis taip pat nėra tinkamas.

Toliau darėme prielaidą, kad savo prognozes turėtumėme apriboti 100%, nes, jei duomenys yra surinkti apklausiant namų ūkius, tikėtina, kad daugiau kaip 100% namų ūkių negali turėti kompiuterių. Pirmiausia taikėme logistinį modelį, iš tiesų jo prognozės daugmaž atitiko lūkesčius, išskyrus dviems šalims: Kinijai ir Estijai. Tačiau yra vienas šio modelio trūkumas, jis yra simetrinis, o nagrinėjamu atveju nėra aišku ar kompiuterių įsigyjimai lėtės tokiu pat tempu kaip kad greitėjo iš pradžių.

Mėginome pritaikyti plačiai šioje srityje naudojamą Gompertz modelį. Nagrinėjome keturis jo variantus. Pirmiausia nagrinėjome Gompertz modelį, kuris rėmėsi ta pačia prielaida, kaip ir logistinis, kad viršutinė augimo asimptotė turėtų būti 100%. Vėliau lubas mėginome didinti atitinkamai iki 120% ir iki 140%. Paskutinį modelį vertinome be jokių apribojimų luboms. Iš tiesų lubų parinkimas yra sudėtingas, juolab, kad tiksliai nežinome, kaip duomenys buvo surinkti. Modelis be apribojimų luboms akivaizdžiai pervertino prognozes. Tačiau modeliai su lubomis iki 120 ir iki 140 prognozavo pakankamai gerai, jų paklaidos buvo pačios mažiausios lyginant su logistiniu bei Gompertz modeliu su lubomis 100%.

Taigi, geriausio modelio parinkimas iš tiesų priklauso nuo prielaidų, kurias darome ir nuo duomenų surinkimo būdo. Būtų įdomu rezultatus palyginti su Bass modeliu, kurio trumpas aprašymas pateiktas darbo pabaigoje ir nors jo įgyvendinimas yra sudėtingas, tačiau turėtų duoti gerus rezultatus.

6. Literatūros sąrašas

1. Jeffrey M. Wooldridge. *Econometric Analysis of Cross Section and Panel Data*. The MIT Press.
2. Dennis Halcoussis. *Understanding Econometrics*. Thomson. JAV (2005).
3. A.C. Cameron. P.K. Trivedi. *Microeconometrics. Methods of Applications*.
4. Nigel Meade. Towhidul Islam. *Forecasting with growth curves: An empirical comparison*. International Journal of Forecasting **11**. 199-215 (1995).
5. Debabrata Talukdar. K. Sudhir. Andrew Ainslie. *Investigating New Product Diffusion Across Products and Countries* **21**. 97-114 (2002)
6. Towhidul Islam. Denzil G. Fiebig. Nigel Meade. *Modelling multinational telecommunications demand with limited data*. International Journal of Forecasting **18**. 605-624 (2002).
7. G. Jarne, J. Sanchez-Choliz, Fatas-Villafranca, S-shaped economic dynamics. The logistic and Gompertz curves generalized, **1048** (2005)
8. V. Mahajan. E. Muller. F. Bass. New product diffusion models in marketing: A review and directions for research. Journal of Marketing **54**. 1-26 (1990).