Keskustelualoitteita #67
Joensuun yliopisto, Taloustieteet

Plant and forest region level approach to forecasting forest chips energy production in Finland

Mikael Linden

ISBN 978-952-219-318-6
ISSN 1795-7885
no 67

PLANT AND FOREST REGION LEVEL APPROACH TO FORECASTING FOREST CHIPS ENERGY PRODUCTION IN FINLAND

Mikael Linden *)

University of Joensuu
Department of Economics and Business
(mika.linden@joensuu.fi)

NOVEMBER 2009

Abstract

Energy policy measures target to increase energy production of forest chips in Finland to 10TWh by year 2010. However at forest region level the production differences are large. Also the regional potential estimates of raw materials base for forest chips production are heterogeneous. In order to analyse the validity of target different methods are proposed to derive forecasts for region level forest chips energy production in Finland in years 2008 - 2014. The plant level data from years 2003 - 2007 gives a starting point for a detailed statistical analysis of present and future region level forest chips production in Finland. Observed year 2008 regional levels are above the estimated prediction 5% confidence intervals based on aggregation of plant level time averages. A simple time trend model with region fixed effects provides accurate forecasts for years 2008 – 2014. Forest chips production forecast confidence intervals cover almost in all regions the year 2008 levels and the potential estimates by year 2014. The forecast confidence intervals are also derived with re-sampling methods, i.e. with bootstrap methods, to obtain more reliable results. Results confirm that a general materials shortcoming is not expected in near future for forest chips energy production in Finland.

*) The author thanks for valuable comments by Jussi Uusivuori and Matti Mäkelä on earlier draft of paper.

1. INTRODUCTION

Bio-energy is part of agenda to mitigate climate change and to reduce the energy dependence on fossil fuels (Faaij 2006). To increase the use of forest biomass is also one of the main targets of EU's energy policy (Parikka 2003). In Finland bio-energy option is strongly connected to country's large forest resources and their industrial use. Wood-based fuels like forest and industrial chips, small tree harvest, pellets, bark, and sawdust are widely used by heating and power plants in Finland (Hakkila 2006). When including also the fuel wood consumption in small-sized dwellings, the share of wood-based fuels in total energy consumption in Finland has increased from 7.9% in 1980 to 9.8% in 2007 (Energy Statistic 2009). The steady increase is expected to accelerate in the 2010's as energy policy is strongly targeted – both in EU and in Finland - to reduce the threat of global warming. The current decade has seen also a growing interest in forest chips as energy source. This is an outcome of new policy orientation with tax reductions, large input materials supplies, and energy production substitution based on fuel input prices (Hakkila 2006). The abundance of residues of forest fellings and thinning like stumps, branches, and small trees builds up a large raw material base for forest chips energy production (VTT 2000). In year 2007 Pöyry Consulting Inc. conducted a research concerning the potential material base of forest chips production in different forest regions. The results implied that large potential supplies remain exploited across the forest regions. Relating the Pöyry estimates to the database of Finnish Forest Research Institute (METLA) that includes data on forest chips energy production and material demands at plant level gives a base for detailed statistical analysis of present and future forest chips energy production in Finland. However price data on forest chips and on its material base is very imperfect hindering estimation of any detailed econometric models.

In the following the target is to take advantage of these data sets in deriving forecasts of forest chips energy production in Finland in coming years. We propose three related methods. Our first method uses efficiently the plant level data in different forest regions. We derive region level aggregate estimates for forest chips material demands by treating plant level inputs as random variables allowing for correlations between input categories and plants. By aggregation of the time averages of plant level production in years 2003 – 2007 we derive region level estimates. The standard errors of estimates are also corrected for plant level autocorrelations. The region level 90% confidence intervals (CI's) are derived to facilitate the statistical analysis of difference between the observed year 2008 regional forest chips energy production and derived estimates. Also the comparison between the potential estimates by Pöyry Inc. and derived estimates is conducted.

The second approach uses simple regression framework to model the material demands as function of forest chips production. We introduce the method of inverse regression (for more details, see van Belle 2002, and references therein) that allows us to analyse how different inputs contribute to forest chip energy production. The method needs to derive standard error of forecasts with the delta method. Although the method is quite case sensitive it provides in this case region level forest chips predictions with some accuracy.

Finally we forecast the region level chips energy production with fixed effect (FE) panel data model where linear time trend acts as forecasting variable. Trend models have a long history in economic forecasting and it works well in short run forecasting (Diebold 1998, Granger 1989). Note that FE-model entails that different regions have region specific level terms but they have common slope estimate, i.e. time trend. Baltagi (2008) shows that slope homogenous parameter models forecast generally better than models where we allow the slope estimate to vary across the regions. The trend forecasts with 90% CI's up till year 2014 are derived under the normality assumption. However this assumption is still questionable for many reasons (see McCullough 1996, Stine 1985) although we do not use stochastic forecaster. In our context the non-normality of residuals in FE trend panel model undermines the derived 90% CI's of forecasts. In order to obtain more reasonable CI's we use re-sampling methods (bootstrap). Re-sampling of model residuals is a convenient method to derive the empirical forecast distributions with panel model having fixed variables (Lam & Veall 2002, Peters & Freedman 1985).

2. FOREST CHIPS OBSERVATIONS

Consider case where energy plant's energy production H_{it} measured in MWh is based on forest chips. In practice the material base of forest chips compromises of two raw material input classes: R_{it} = stumps and residues from forest fellings, and S_{it} = small tree harvest. Because the measurement of these inputs is in same energy content (MWh) as H_{it}, the identity $H_{it} \equiv R_{it} + S_{it}$ holds for all plants $i = 1,...,N$ at any time point $t = 1,..,T$. However the shares of R_{it} and S_{it} vary across the plants. Likewise shares do not remain constant across the time for individual plants.

We derive forecasts for forest chips energy production at forest region level based on plant level production and demand observations in years 2003 - 2007. The data is provided by Finnish Forest Research Institute (METLA). The coverage of plants is almost 100%. However our panel data is unbalanced, i.e. many plants do not produce energy with forest chips in all sample years. The number of plants that has a positive value of forest chips energy production at least for one year in the sample is 417. The derived region level forecasts are compared with the forest region potential supplies of R_g^P and S_g^P. These estimates are provided by Pöyry Consulting Inc. (2008) for year 2007. We also compare forecasts to the observed - year 2008 - forest chips region production levels. The evident trend in region 2003 - 2008 observations supports the proposed trend model as working horse for post sample forecasting.

3. PLANT LEVEL APPROACH: TIME MEAN PREDICTIONS

Assume that the forest chips energy output of each separate plant is Normal distributed with plant specific finite expectation and variance. Notice also that each plant is located in define forest region *g =1,...,M.* Now

1) $$H_{igt} = R_{igt} + S_{igt}, \quad \text{where } i = 1,\ldots,N_g,\ \ t = 1,\ldots,T,\ \text{ and }\ g = 1,\ldots,M.$$ [1]

2) $$\begin{bmatrix} R_{itg} \\ S_{itg} \end{bmatrix} \sim N\left[\begin{pmatrix} \rho_{ig} \\ \sigma_{ig} \end{pmatrix}, \begin{pmatrix} \theta^2_{R,ig} & \theta_{RS,ig} \\ \theta_{SR,ig} & \theta^2_{S,ig} \end{pmatrix}\right], \quad \text{for all } t = 1,\ldots,T.$$

The model entails that we can estimate the time averages of each plant's demand for stumps and residues $\overline{R}_{ig}$, and small tree harvest $\overline{S}_{ig}$ for time period 2003 - 2007. We also obtain easily the standard errors (SE) of time mean estimates.[2]

3) $$\overline{R}_{ig} = \frac{1}{T}\sum\nolimits_{t=1}^{T} R_{itg}, \qquad \mathrm{SE}(\overline{R}_{ig}) = \sqrt{VAR[\overline{R}_{ig}]},\ \text{ where } VAR[\overline{R}_{ig}] = \hat{\theta}^2_{R,ig}\,/\,T,$$

$$\overline{S}_{ig} = \frac{1}{T}\sum\nolimits_{t=1}^{T} S_{itg}, \qquad \mathrm{SE}(\overline{S}_{ig}) = \sqrt{VAR[\overline{S}_{ig}]},\ \text{ where } VAR[\overline{S}_{ig}] = \hat{\theta}^2_{S,ig}\,/\,T.$$

---

[1] To keep the notation simply we do not write down the time indexes for unbalanced data, i.e. $t_i = 1,..,T_i$.

[2] Note that variances can be harmed by plant level autocorrelation. E.g. if observations X_t follow *AR(1)* process with autocorrelation coefficient $\rho > 0$, then the true standard error of $\overline{X}_t$ is $\mathrm{SE}(\overline{X}_t) \approx \sqrt{(1+\rho)/(1-\rho)}\mathrm{std}(X_t)/\sqrt{T}$ $> \mathrm{std}(X_t)/\sqrt{T}$. Note also that $E[\overline{X}_t] = \mu/(1-\rho)$.

$\hat{\theta}^2_{R,ig}$ and $\hat{\theta}^2_{S,ig}$ are the variance estimates of $R_{t,ig}$ and $S_{t,ig}$. Next we aggregate the estimated plant level time means to forest region level, i.e. we get an estimate for total forest region level demand for R_g and S_g, for each $g = 1,...,M$

$$R_g = \sum_{i=1}^{N_g} \bar{R}_{ig} = \frac{1}{T}\sum_{t=1}^{T} R_{1tg} + ... + \frac{1}{T}\sum_{t=1}^{T} R_{N_g tg}$$

$$S_g = \sum_{i=1}^{N_g} \bar{S}_{ig} = \frac{1}{T}\sum_{t=1}^{T} S_{1tg} + ... + \frac{1}{T}\sum_{t=1}^{T} S_{N_g tg} .$$

This means that at forest region level the total forest chips production is estimated as a sum of R_g and S_g,

$$H_g = \sum_{i=1}^{N_g} \bar{H}_{ig} = R_g + S_g = \sum_{i=1}^{N_g} \bar{R}_{ig} + \sum_{i=1}^{N_g} \bar{S}_{ig} , \quad \text{for all } g = 1,...,M.$$

Because each component in the sums is random we can derive variances for R_g and S_g. Notice that we allowed for correlation between plant level R_{ig} and S_{ig} for each plant i in given forest region *g*. Thus the covariance terms $\theta_{RS,ig}$ are also present in the analysis

$$VAR[H_g] = VAR[R_g] + VAR[S_g] + COV[R_g, S_g]$$

5) $$= VAR[\sum_{i=1}^{N_g} \bar{R}_{ig}] + VAR[\sum_{i=1}^{N_g} \bar{S}_{ig}] + COV[\sum_{i=1}^{N_g} \bar{R}_{ig}, \sum_{i=1}^{N_g} \bar{S}_{ig}]$$

$$= \sum_{i=1}^{N_g} \frac{\hat{\theta}^2_{R,ig}}{T} + \sum_{i=1}^{N_g} \frac{\hat{\theta}^2_{S,ig}}{T} + \frac{1}{T}\sum_{i=1}^{N_g} \hat{\theta}_{RS,ig} .$$

We observe also that forest chips energy production of plants in a given region are likely correlated with each other as the plants competitive for the same input materials. We expect the correlations to be negative. An estimate for cross-section dependence between the plants in a given forest region is derived. The structure of data, unbalanced panel, makes the derivation of covariance estimate difficult since we have only in some cases all observations for years 2003 – 2007, i.e. T = 5. Next we use only these observations to derive estimate $\hat{\theta}_g$, i.e. for $n_g < N_g$

$$\hat{\theta}_g = \sum_{i=1}^{ng-1} \sum_{j=1}^{ng} \hat{\theta}_{ij,g} \text{ , where } \hat{\theta}_{ij,g} = \frac{1}{T-1} \sum_{t=1}^{T} (H_{it,g} - \bar{H}_{it,g})(H_{jt,g} - \bar{H}_{jt,g}) .$$

Finally the 90% approximate normal confidence intervals (CI's) for each H_g, $g = 1,...,M$ are

6) $$H_g \pm 1.65 \sqrt{\frac{1}{T} \sum_{i=1}^{N_g} \hat{\theta}_{R,ig}^2 + \frac{1}{T} \sum_{i=1}^{N_g} \hat{\theta}_{S,ig}^2 + \frac{1}{T} \sum_{i=1}^{N_g} \hat{\theta}_{RS,ig} + \hat{\theta}_g} .$$

The result enables to compare to what extension the observed year 2008 values $H_{g,2008}$ and $S_g^P + R_g^P = H_g^P$, the potential supply estimates of Pöyry Consulting Inc. are inside the derived 90% CI's. Thus, if we able to show that year 2008 production and potential forest chips resources are above the upper 95% CI-level of 2003 – 2007 plant average production, we can inference that there is a statistically significant difference between region levels of chips forest energy production compared to the potential levels and year 2008 observed levels. Conversely, if year 2008 region production levels or potential resources are inside the 90% CI's, we inference that there is statistically no difference between current or potentially production levels compared to year 2003 – 2007 region levels.

We observe that in all regions, expect for region 1, the potential estimates by Pöyry Inc. H_g^P are above the upper 95% margin. In region 1 the Pöyry estimate is below the 95%-LOW margin. Similarly the actual year 2008 levels of forest chips energy production are for all regions, except for region 5, above the mean 2003-2007 values, $\bar{H}_g^{2003-2007}$. These results are expected since the estimated 90% CI's are quite narrow. However they are evidently biased since they are based on assumption of Normal distribution. Also all firm specific autocorrelations and some cross-section correlations are neglected in analysis. Likewise potential forest resource estimates are prone to measurement errors. Irrespectively of these inference problems the region forest chips production outputs have increased from average 2003 –2007 levels. The figure 1 in Appendix I gives year 2003 index levels of forest chips energy production in different forest regions in years 2003 – 2008. There exists a clear trend upwards in chips production in almost all regions.

TABLE 1. 90% CONFIDENCE INTERVALS OF FOREST CHIPS ENERGY PRODUCTION IN DIFFERENT FOREST REGIONS, $\bar{H}_g^{2003-2007}$

Region	95%-LOW	$\bar{H}_g^{2003-2007}$	95%-HIGH	H_g^P	$H_{g,2008}$
1	554.84	706.86	858.87	***548.00***	**1376.14**
2	456.17	530.08	603.98	**1105.40**	**608.52**
3	368.61	440.48	512.35	**1526.30**	**678.16**
4	432.76	493.28	553.81	**1131.60**	**610.01**
5	309.54	356.89	404.24	**1370.10**	400.51
6	390.49	462.92	535.36	**1774.80**	**740.82**
7	210.27	254.79	299.31	**474.80**	**401.09**
8	712.39	820.54	928.69	**1145.10**	**1181.80**
9	147.67	186.00	224.32	**1429.90**	**378.99**
10	315.71	364.98	414.24	**1415.10**	**553.65**
11	88.46	196.35	304.24	**792.60**	**499.05**
12	220.24	261.17	302.10	**643.40**	**403.40**
13	119.02	141.45	163.87	**745.80**	**207.65**

3. INVERSE REGRESSION –APPROACH

Notice that our interest lies in the representative plant presentation of forest chips resource demand at forest region level. This observation leads to following panel data model as a starting point for modelling plants' forest chips material demands in forest district *g*

7A) $$R_{it} = \alpha_1 + \beta_1 H_{it} + \varepsilon_{1,it} \quad | R_{it} > 0 \quad \text{for } \forall it .$$

7B) $$S_{it} = \alpha_2 + \beta_2 H_{it} + \varepsilon_{2,it} \quad | S_{it} > 0 \quad \text{for } \forall it$$

Note when all plants demand both inputs at same time (i.e. both R_{it} and S_{it} are positive for each plant), the above equations are mirror images of each other. That is

$$R_{it} + S_{it} = (\alpha_1 + \alpha_2) + (\beta_1 + \beta_2) H_{it} + \varepsilon_{1,it} + \varepsilon_{2,it} = H_{it} \quad \text{for } \forall it .$$

However this approach is too restricted for a proper forecasting analysis. It throws away observations since many plants use only one material input. Thus the two equation model above has its justification because the restriction $\beta_1 + \beta_2 = 1$ holds for firms that have both $R_{it} > 0$ and $S_{it} > 0$. An alternative approach would be a Tobit type approach where we allow for $R_{it} = 0$ and $S_{it} = 0$ observations in their "demand" equations. However it is hard to find interpretation for this approach. Also the distribution assumptions of Tobit model are hardly met in our data. Instead we defence the approach above where $R_{it} = 0$ or $S_{it} = 0$ observations are excluded from analysis without biasing LS-estimates upward. This leads to different sample sizes for each demand equation since plants use heterogeneously only $R_{it} > 0$ or $S_{it} > 0$ across the sample period.

As our target is to forecast forest chips energy production at the forest region level we can use model 7A) –7B) to derive $H_{g,T+1}$ forecast with the *inverse regression* approach. The method is easy and natural one is this context. Note that OLS estimation of 7A)

$$R_{it} = \alpha_1 + \beta_1 H_{it} + \varepsilon_{1,it} \quad \text{with} \quad \varepsilon_{1,it} \sim IID(0, \sigma^2_{1,\varepsilon}) \quad (i = 1,..., N_g)$$

results to

$$8) \qquad \hat{R}_{it} = \hat{\alpha}_1 + \hat{\beta}_1 H_{it} \quad \Rightarrow \quad H^{est}_{1,it} = \frac{R^{obs}_{it} - \hat{\alpha}_1}{\hat{\beta}_1},$$

where R^{obs}_{it} is some representative observation for R_{it}. The variance of $H^{est}_{1,it}$ using delta method is made up of four terms:

$$9) \qquad VAR[\hat{H}^{est}_{1,it}] \approx \frac{\hat{\sigma}^2_{1,\varepsilon}}{\hat{\beta}^2_1} + \frac{VAR[\hat{\alpha}_1]}{\hat{\beta}^2_1} + (\frac{R^{obs}_{it} - \hat{\alpha}_1}{\hat{\beta}^2_1})^2 VAR[\hat{\beta}_1] + 2(\frac{R^{obs}_{it} - \hat{\alpha}_1}{\hat{\beta}^3_1}) COV[\hat{\alpha}_1, \hat{\beta}_1].$$

Similar derivation is valid also for model 7B), i.e. $\hat{S}_{it} = \hat{\alpha}_2 + \hat{\beta}_2 H_{it} \quad \Rightarrow \quad H^{est}_{2,it} = \dfrac{S^{obs}_{it} - \hat{\alpha}_2}{\hat{\beta}_2}$.

Next we estimate equations 7A) and 7B) with OLS for each forest region and "invert" them using $R^{obs}_{it} = \overline{R}_g$ and $S^{obs}_{it} = \overline{S}_g$ to derive forest region level forecasts $\hat{H}_g$. Note that we get two forecasts for H_g i.e. $\hat{H}^{est}_{1,g}$ and $\hat{H}^{est}_{2,g}$. This enables us to evaluate separately the forecasts performance of $\overline{R}_g$ and $\overline{S}_g$ compared to potential forest chips resources and year 2008 observed production.

TABLE 2. INVERSE OLS 90% CONFIDENCE INTERVALS OF FOREST CHIPS ENERGY PRODUCTION AT DIFFERENT FOREST REGIONS

Region		95%-LOW	$H_{1,g}^{est} / H_{2,g}^{est}$	95%-HIGH	H_g^P	$H_{g,2008}$
1	$\overline{R}_g$	619.87	642.04	664.22	**548.00**	**1376.14**
	$\overline{S}_g$	914.42	1038.89	1163.36	**548.00**	**1376.14**
2	$\overline{R}_g$	427.17	439.24	451.32	**1105.40**	**608.52**
	$\overline{S}_g$	1131.17	1559.02	1986.87	**1105.40**	**608.52**
3	$\overline{R}_g$	339.98	363.93	387.89	**1526.30**	**678.16**
	$\overline{S}_g$	547.48	695.37	843.27	**1526.30**	678.16
4	$\overline{R}_g$	369.26	402.27	435.27	**1131.60**	**610.01**
	$\overline{S}_g$	238.77	666.57	1094.37	**1131.60**	610.01
5	$\overline{R}_g$	289.79	302.04	314.29	**1370.10**	**400.51**
	$\overline{S}_g$	776.41	1301.42	1826.42	1370.10	**400.51**
6	$\overline{R}_g$	298.99	319.95	340.91	**1774.80**	**740.82**
	$\overline{S}_g$	668.40	711.65	754.90	**1774.80**	740.82
7	$\overline{R}_g$	101.68	119.67	137.66	**474.80**	**401.09**
	$\overline{S}_g$	329.26	386.39	443.52	474.80	**401.09**
8	$\overline{R}_g$	744.19	759.45	774.71	**1145.10**	**1181.80**
	$\overline{S}_g$	1448.82	1952.28	2455.73	**1145.10**	**1181.80**
9	$\overline{R}_g$	114.94	121.28	127.62	**1429.90**	**378.99**
	$\overline{S}_g$	81.10	93.32	105.54	**1429.90**	**378.99**
10	$\overline{R}_g$	156.45	175.05	193.65	**1415.10**	**553.65**
	$\overline{S}_g$	623.87	958.50	1293.13	**1415.10**	**553.65**
11	$\overline{R}_g$	120.07	140.83	161.60	**792.60**	**499.05**
	$\overline{S}_g$	251.79	546.13	840.46	792.60	499.05
12	$\overline{R}_g$	123.03	139.05	155.06	**643.40**	**403.40**
	$\overline{S}_g$	571.05	667.51	763.98	643.40	**403.40**
13	$\overline{R}_g$	45.29	78.09	110.90	**745.80**	**207.65**
	$\overline{S}_g$	101.74	107.43	113.12	**745.80**	**207.65**

Table 2 reports the results from inverted OLS method. Although the method gives very narrow forecast 90% -intervals, the forecasts with $\overline{R}_g$ are too low. Contrary to this the forecasts with $\overline{S}_g$

are too high. The result is expected since the forecasts depend on the estimated parameter values of basic “demand” equations 7A) and 7B). The parameter β_1 gets value close to one and β_2 is close to zero (see Appendix II). Note that all cases where potential estimates or observed year 2008 values are inside the 90% CI’s are obtained with $\overline{S}_g$. Is this sense small tree demand model 7B) and it’s inverted forecast formula provide more valid point forecasts than the similar model for $\overline{R}_g$. Note that broader 90% CI’s would have been obtained if corrections for aggregation and correlations in $VAR[H_{1,it}^{est}]$ and $VAR[H_{2,it}^{est}]$ had been considered.

4. FOREST REGION LEVEL APPROACH WITH TREND –MODEL

The focus of our analysis is next on the region level forest chips production in years 2003 – 2014. Aggregated data consist of sums of plant level forest chips energy outputs in different forest regions $g = 1,...,13$ in years 2003 - 2007. Thus our data set is now balanced panel with size of $N \times T = 13 \times 5 = 65$. We change the analysis to region level because of the plant level heterogeneity, i.e. few time observations per plant cause large dispersions to fixed effects (FE) model estimates at plant level. Note also that the potential and observed year 2008 forest chips outputs are recorder only at forest sector level. The aggregation from plant level to region level is prone to some uncertainty (see above) but the aggregated values are unbiased.

Under assumption that forest chips energy demand, like any other bio-fuel, will increase in coming next years, a panel data trend model like

$$10) \qquad H_{g,t} = a_g + \beta t + \varepsilon_{g,t}, \quad \text{with } \varepsilon_{g,t} \sim NID(0, \sigma_\varepsilon^2) \text{ for } g = 1,...,M \quad \text{and } t = 0,1,...,T$$

is a convenient model alternative in this context, since forecasts are derived easily with post-sample observation $t*$ where $t* = T + 1 = 5$ as $2003 = 0,\ 2004 = 1,...,2007 = 4 = T$. Assume that we have an estimate of trend model for region based observations of $H_{g,t}$ in ex-ante sense $t* \le T$

$$H_{g,t} = \hat{\alpha}_g + \hat{\beta} t + \hat{\varepsilon}_{g,t}.$$

The result is used to derive the following empirical forecasts or predictions for $g = 1,..,M$

$$\hat{H}_{g,t^*} = \hat{\alpha}_g + \hat{\beta}t^* \qquad | \ t^* = T+1$$

for the true, ex-post observed $t^* > T$, region level state

11) $$H_{g,t^*} = \alpha_g + \beta t^* + \varepsilon_{g,t^*}.$$

The forecast error is defined as

12) $$e_{g,t^*} = H_{g,t^*} - \hat{H}_{g,t^*} = (\alpha_g - \hat{\alpha}_g) + (\beta - \hat{\beta})t^* + \varepsilon_{g,t^*}.$$

Note that $E[\varepsilon_{g,t^*}] = 0$. Therefore FE–forecasts are unbiased. In general, the forecasts are subject at least to four different sources of error. First, since we have only sample estimates of α_g and β, one source is the sampling error of parameter estimates. The second is the error term of forecast ε_{g,t^*} that always will be present. The third source of error or uncertainty - not present here - is the fact that predictor is usually also random. In addition we typically observe predictors with measurement errors. Lastly we can argue that there exists also parameter uncertainty concerning the model, i.e. the parameters α_i and β are a priori subject to random variation. In this context we pay attention only to the first two sources of error.

The variance of the forecast error is

13) $$VAR[e_{g,t^*}] = VAR[\hat{\alpha}_g] + COV[\hat{\alpha}_g, \hat{\alpha}_h] + t^{*2}\, VAR[\hat{\beta}] + 2t^*\, COV[\hat{\alpha}_g, \hat{\beta}] + \sigma^2_{\varepsilon_{g,t^*}}.$$

If FE-panel estimation without overall intercept is conducted, then the balanced panel data FE_{OLS}-model in stacked vector form is

$$\boldsymbol{y} = [\boldsymbol{d}_1 \ \boldsymbol{d}_2 \ .... \ \boldsymbol{d}_N \quad \boldsymbol{X}]\begin{bmatrix}\boldsymbol{\alpha}\\ \boldsymbol{\beta}\end{bmatrix} + \boldsymbol{\varepsilon} = \boldsymbol{D\alpha} + \boldsymbol{X\beta} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{d}_i$'s are $N \times T$ vectors containing T length sub-vector of ones collected into matrix $\boldsymbol{D}_{(NT \times N)}$. Different prediction error components have following estimates, where $\bar{\boldsymbol{X}}_{g\cdot}$ is the mean of *t = 0,1,…,T*,

$$VAR[\hat{\alpha}_g] = \frac{\hat{\sigma}_\varepsilon^2}{T} + \bar{\boldsymbol{X}}_{g\cdot}^{'} VAR[\hat{\beta}]\bar{\boldsymbol{X}}_{g\cdot}, \qquad COV[\hat{\alpha}_g, \hat{\alpha}_h] = \bar{\boldsymbol{X}}_{g\cdot}^{'} VAR[\hat{\beta}]\bar{\boldsymbol{X}}_{g\cdot},$$

$$COV[\hat{\alpha}_g, \hat{\beta}] = -\bar{\boldsymbol{X}}_{g\cdot}^{'} VAR[\hat{\beta}], \qquad VAR[\hat{\beta}] = \hat{\sigma}_\varepsilon^2 / VAR(t),$$

$$\hat{\sigma}_\varepsilon^2 = \sum_{g=1}^{N} \sum_{t=1}^{T+1} \hat{\varepsilon}_{g,t}^2 / [T_N - N - 1], \quad \text{where } T_N = N \times (T+1).$$

Note that we do not derive variance estimate for error of forecast $\sigma_{\varepsilon_{g,t^*}}^2$ since we have only one observation for it, i.e. $e_{g,t^*}^{obs} = H_{g,t^*}^{obs} - \hat{H}_{g,t^*}$. An efficient estimator for $\sigma_{\varepsilon_{g,t^*}}^2$ is $\hat{\sigma}_\varepsilon^2$. Sometimes $\hat{\sigma}_{\varepsilon_g}^2$ is also used, but this line is not pursued here because we have only 5 time observations for each region.

Finally we observe that 90% approximate forecast confidence interval (CI) for $\hat{H}_{g,T+1}$ under normality is

$$14) \qquad [\hat{\alpha}_g + \hat{\beta}(T+1)] \pm 1.65\sqrt{VAR(\hat{e}_{g,t^*})}.$$

Year 2008 forecasts

For period $t^* = T + 1 = 5$ (year 2008) we use estimated fixed effects $\hat{\alpha}_g$ as base forecasts when deriving forecasted $\hat{H}_{g,T+1}$ values. The argument is that level of region forest chips energy production starts from year *t = 0* (year 2003) level and all production growth is captured by trend estimate. Error of this assumption is captured by $VAR[\hat{\alpha}_g]$. Thus the forecasts error variance and 90% CI's are derived with Eqs. 13) and 14). Note that Eq. 13) entails many error components. FE_{OLS}-estimation results in Appendix II show that residuals are non-normal but not auto-correlated. Despite these problems Table 3 gives reasonable forecast values for year 2008 forest chips energy production. Trend model estimates capture year 2008 region energy output levels inside the 90% CI's in all regions, except in region 1.

TABLE 3. 90% CONFIDENCE INTERVALS OF TREND MODEL FORECASTS FOR FOREST CHIPS ENERGY PRODUCTION IN 2008 WITH ESTIMATED FOREST REGION LEVELS

Region	95%-LOW	$\hat{H}_{g,2008}$	95%-HIGH	$H_{g,2008}$	Forecast error *)
1	558.28	866.38	1174.48	**1376.14**	509.76 (37%)
2	330.24	638.34	946.44	608.52	-29.82 (-5%)
3	249.56	557.66	865.76	678.16	120.49 (18%)
4	343.62	651.72	959.82	610.01	-41.72 (-7%)
5	170.89	477.99	785.09	400.51	-77.72 (-19%)
6	351.58	659.68	967.78	740.82	81.13 (11%)
7	81.64	389.74	697.65	401.09	11.34 (3%)
8	636.91	944.01	1251.91	1181.80	237.80 (20%)
9	6.81	314.72	622.63	378.99	64.27 (17%)
10	212.09	520.00	827.91	553.65	33.65 (6%)
11	60.55	368.46	676.56	499.05	130.59 (26%)
12	89.64	397.54	705.45	403.40	5.86 (1%)
13	-25.27	282.63	590.54	207.65	-74.98 (-36%)

*) %-values refer to $100 * \frac{(H_{g,2008} - \hat{H}_{g,2008})}{H_{g,2008}}$

In order to evaluate correctness of CI's based on normality assumption in Eq. 14) we derive the forecasts confidence intervals also with re-sampling methods, i.e. by bootstrap methods. In this context bootstrap is easily conducted. We first estimate the trend model in Eq. 11) and obtain the fit $H_{g,t} = \hat{\alpha}_g + \hat{\beta}t + \hat{\varepsilon}_{g,t}$. Secondly, we re-sample randomly the forest region specific residuals $\hat{\varepsilon}_{g,t}$ separately to obtain new residuals $\hat{\varepsilon}^*_{g,t}$. Next we derive new values for forest chips observations with $H^*_{g,t} = \hat{\alpha}_g + \hat{\beta}t + \hat{\varepsilon}^*_{g,t}$. New $H^*_{g,t}$ values enable us to derive new FE_{OLS} –estimates for α_g and β with given fixed trend observations. Repeating this re-sampling procedure 10.000 times with each time also deriving the year 2008 forecasts $\hat{H}_{g,t*} = \hat{\alpha}_g + \hat{\beta}t*$ $(t* = T+1)$ leads to 10.000 distinct estimates for $\hat{H}_{g,t*}$. Their empirical distribution gives 90% CI's. Note that in this procedure we do not estimate any variances or covariances. Also the forecast errors are not explicitly derived. The random re-sampling of $\hat{\varepsilon}_{g,t}$, and the corresponding derived (random) estimates for α_g and β giving the distribution of $\hat{H}_{g,t*}$ estimates, conducts now as basis of analysis of forecast uncertainty.

Thus we are deriving the sampling distribution of conditional mean of forecasts for each region. If we add the $t^* = T + 1$ re-sampled residuals $\hat{\varepsilon}^*_{g,t^*}$ to forecasts we obtain the distribution of conditional forecasts (for more details, see McCullough 1996). We will report the latter ones.

The bootstrap results in Table 4. reveals that CI's are now in many regions much narrow than in Table 3. However the median values of bootstrapped forecasts, $\hat{H}^{MED}_{g,2008}$, are very close to $\hat{H}_{g,2008}$ values in Table 3. confirming us that the distributions of conditional forecasts are symmetric. The observed $H_{g,2008}$ values are in Table 5. in four regions above the upper 95% CI.

TABLE 4. BOOTSTRAP 90% CONFIDENCE INTERVALS OF TREND MODEL FORECASTS FOR FOREST CHIPS ENERGY PRPODUCTION IN 2008 WITH ESTIMATED FOREST REGION LEVELS

Region	95%-LOW	$\hat{H}^{MED}_{g,2008}$	95%-HIGH	$H_{g,2008}$
1	574.87	866.14	1143.30	**1376.14**
2	567.08	638.29	707.25	608.52
3	498.08	558.07	615.59	**678.16**
4	483.25	654.30	800.65	610.01
5	379.21	478.70	577.57	400.51
6	417.21	651.91	964.18	740.82
7	308.11	390.21	466.11	401.09
8	532.99	953.27	1287.62	1181.80
9	247.43	314.76	381.57	378.99
10	456.21	520.08	583.39	553.65
11	274.24	367.85	465.79	**499.05**
12	339.55	397.41	455.24	403.40
13	173.12	282.34	383.54	207.65

The outcome from Tables 3. and 4. is the result that trend model forecasts quite well the year 2008 forest chips energy production across the forest regions. However the CI's based on normal approximation with many different error sources result in quite broad CI's. More accurate CI's are obtained with re-sampling methods showing that in some regions year 2008 forest chips energy outputs are still outside the upper 95% CI's.

Year 2014 forecasts

Using observed values $H_{g,2008}$ as estimates for region specific fixed effects (α_g's) is an experiment that allows us the see if forecasted forest chips output values in different regions after year 2008 has reached the potential resource levels of H_g^P. The method enables us to analyse how forest chips energy production is expected to grow after 2008. Thus we report next forecast values for year 2014 (i.e. $t^* = T + 6 = 11$) based on observed region values $H_{g,2008}$ as estimates for fixed effects α_g. The estimate for variance of the forecast error is now

$$VAR[\hat{e}_{g,t^*}] = t^{*2}\, VAR[\hat{\beta}] + \hat{\sigma}_{\varepsilon}^2, \qquad \text{13')}$$

since fixed effects are not estimated, i.e. $H_{g,2008}$ values are given constants. Trend values and year 2003 – 2007 estimate for $\hat{\beta}$ give now the forecast errors.

TABLE 5. 90% CONFIDENCE INTERVALS OF TREND MODEL FORECASTS FOR FOREST CHIPS ENERGY PRODUCTION IN 2014 WITH YEAR 2008 FOREST REGION LEVELS

Region	95%-LOW	$\hat{H}_{g,2014}$	95%-HIGH	H_g^P	Forecast error *)	
1	1492.12	1978.53	2464.94	***548.00***	-1431.62	(-261%)
2	724.50	1210.91	1697.32	1105.40	-106.08)	(-10%)
3	794.14	1280.55	1766.96	1526.30	245.91	(16%)
4	725.98	1212.40	1698.81	1131.60	-78.16	(-7%)
5	516.48	1002.90	1489.31	1370.10	370.36	(27%)
6	856.80	1343.21	1829.62	1774.80	430.43	(24%
7	517.07	1003.48	1489.89	***474.80***	-528.81	(-111%)
8	1297.78	1784.19	2270.60	***1145.10***	-639.27	(-56%
9	494.97	981.38	1467.79	1429.90	448.02	(31%)
10	669.63	1156.04	1642.45	1415.10	259.00	(18%)
11	615.03	1101.44	1587.85	792.60	-310.09	(-39%)
12	519.38	1005.79	1492.20	643.40	-363.76	(-56%)
13	323.63	810.04	1296.46	745.80	-61.51	(-8%)

*) %-values refer to $100 * \frac{(H_g^P - \hat{H}_{g,2014})}{H_g^P}$

In Table 5. we observe that in year 2014 the 90% CI's of forecasted forest chips outputs contain in all regions, except for regions 1,7 and 8, the estimated potential levels by Pöyry Consulting Inc. H_g^P. In regions 1, 7 and 8 forecasts "shoot over" the potential estimates. Note that CI's are quite broad as $t^* = 11$ in Eq. 13'). Appendix IV gives more detailed picture of forecasts in years 2008 - 2024 in forest regions. Table 6. shows the bootstrap 90% CI's and median forecast values derived with similar re-sampling methods as above for year 2008 forecasts. The CI's are again narrow compared to the normal approximation CI's. In many regions the potential forest output levels estimated by Pöyry Inc. will not reached. However over-shooting takes again place for regions 1,7 and 8 but also in regions 11 and 12.

TABLE 6. BOOTSTRAP 90% CONFIDENCE INTERVALS OF TREND MODEL FORECASTS FOR FOREST CHIPS ENERGY PRODUCTION IN 2014 WITH YEAR 2008 FOREST REGION LEVELS

Region	95%-LOW	$\hat{H}_{g,2014}^{MED}$	95%-HIGH	H_g^P
1	882.17	1979.62	2207.02	***548.00***
2	1030.04	1211.48	1390.25	1105.40
3	1100.85	1280.39	1459.24	**1526.30**
4	1019.34	1209.76	1405.38	1131.60
5	822.31	999.74	1184.35	**1370.10**
6	1132.75	1344.37	1580.19	**1774.80**
7	817.61	1003.61	1188.21	***474.80***
8	1505.71	1784.37	2025.26	***1145.10***
9	810.52	981.88	1161.52	**1429.90**
10	975.36	1156.10	1159.61	**1415.10**
11	916.76	1102.69	1335.17	***792.60***
12	829.38	1007.16	1279.20	***643.40***
13	624.99	807.31	992.46	745.80

As a summary we observe that independently how we derive the 90% CI's of region specific forecasts the trend model approach gives forecast values that are not of secondary value. Results in Tables 3 – 6 show that in most forest regions year 2008 forest chips energy production levels are forecasted with precision. The potential resource levels are reached in some regions before the year 2014. Note that potential estimates provided by Pöyry Inc. contain measurement errors that should be counted for. For example if we allow 20% error margin for Pöyry estimates (i.e. $H_g^P \pm 0.1 \times H_g^P$)) then almost all forecasted CI's cover the region specific potential estimates in Table 6.

5. CONCLUSIONS

The demand for forest energy is expected increase in Finland during the next decade. Residual forest biomass is abundantly available. The capacity of heating and power plants to use forest chips is large enough to meet the supply. Different policy measures started already in mid 1990's are targeted to increase the production of forest chips in Finland to 10TWh in 2010 (Parikka 2006). The estimate of total current potential forest chips energy content is 14.1TWh (Pöyry Consulting Inc. 2007). Year 2008 forest chips energy production level was 80.4TWh.

These numbers ask for more detailed forecast analysis based on plant and forest region data. The paper proposed different prediction methods to derive region level forest chips forecasts for years 2008 – 2014. The results show that year 2008 region chips energy levels are not predicted with year 2003 – 2007 time average plant level data although confidence intervals are corrected for spatial, resource input, and temporal correlations. The proposed inverse regression method based on forest chips resource input demands resulted also in predictions that not covered year 2008 and potential levels. However in some cases forecasts were also over-shooting asking the validity of inverse method.

A simple trend forecast model based on panel data of forest regions in years 2003 – 2007 provided reasonable predictions for years 2008 – 2014. The observed aggregate - the whole country - production level for year 2008, year 2010 prediction, and potential aggregate are crossed. The region forecast confidence intervals cover almost in all regions the year 2008 chips energy levels and the potential estimates by year 2014. A re-sampling approach was conducted to derive more reliable confidence intervals for region forecasts because residuals of panel data fixed effects model were non-normal. Less coverage was found with re-sampling approach. The potential material base estimates by Pöyry Inc. are in some forest regions still reached but also more over-shooting cases are found. However a general materials shortcoming is not expected in near future. The fact that forest region level results are still heterogeneous demands future research wherein forest region specific dependencies and factors (e.g. prices and transportation costs) are used.

REFERENCES

Baltagi, B.H. (2008) "Forecasting with Panel Data", *Journal of Forecasting* **27**, 153 – 173.

van Belle, G. (2002) *Statistical Rules of Thumb*, Wiley, NY.

Diebold, F.X. (1998) *Elements of Forecasting*, Thomson Int., Ohio.

Energy Statistics (2009) Yearbook 2008, Statistics Finland, Helsinki.

Faaij, A.P.C (2006) "Bio-Energy in Europe: Changing Technology Choices", *Energy Policy* **34**, 322-342.

Granger, C.W.J. (1989) *Forecasting in Business and Economics*, 2nd, Academic Press, NY.

Hakkila, P. (2006) "Factors Driving the Development of Forest Energy in Finland" *Biomass and Energy* **30**, 281-288.

Lam, J.-P. and Veall, M.R. "Bootstrap Prediction Intervals for Single Period Regression Forecasts", *International Journal of Forecasting* **18**, 125-130.

McCullough, B.D. (1996) "Consistent Forecast Intervals when the Forecast-period Exogenous Variables are Stochastic", *Journal of Forecasting* **15**, 293 - 304.

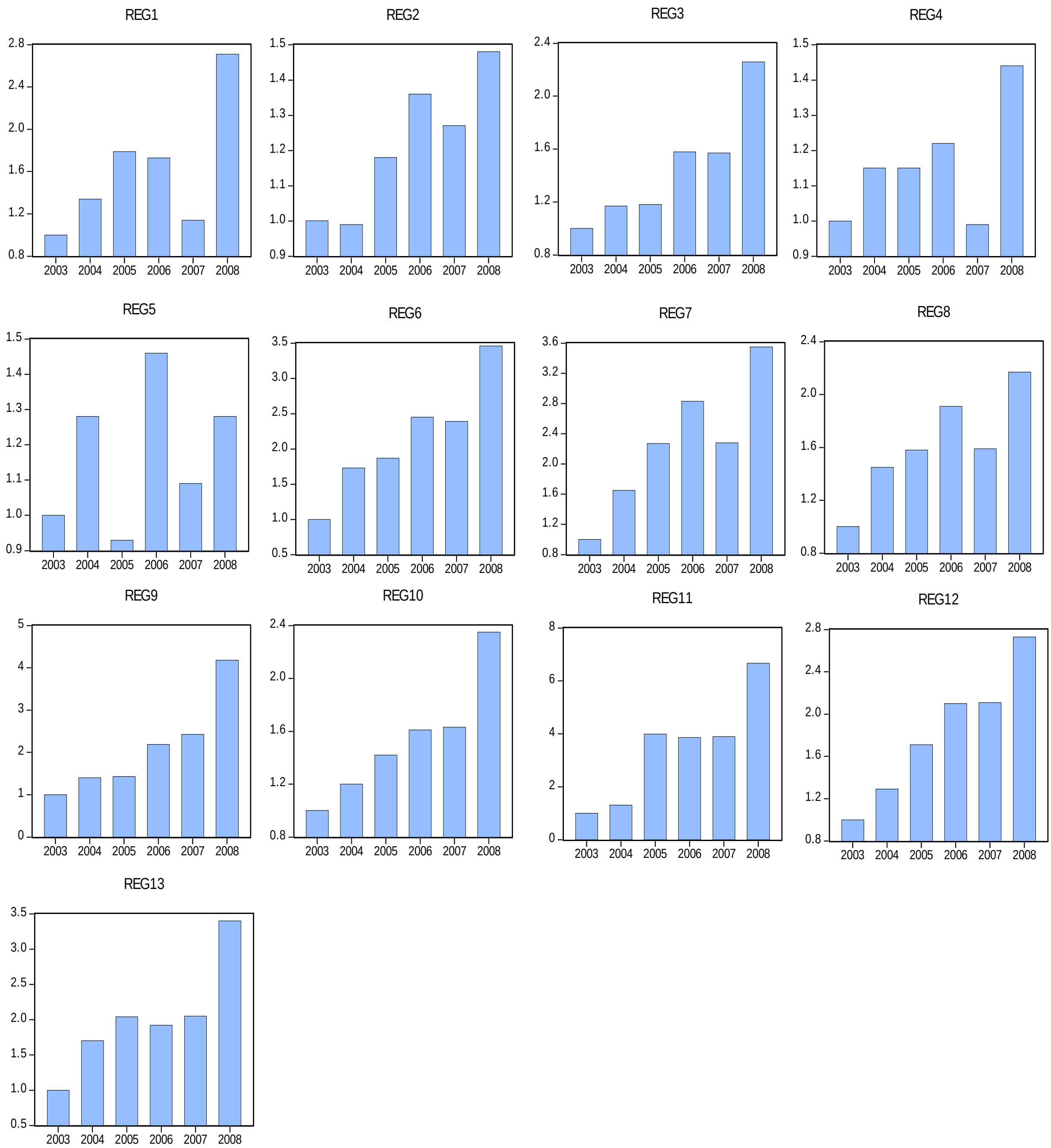
Parikka, M. (2003) "Woody Biomass Resources in Europe", Department of Bioenergy, SLU.

Peters, S.C and Freedman, D.A. (1985) "Using the Bootstrap to Evaluate Forecasting Equations" , *Journal of Forecasting* **4**, 251 – 262.

Stine, R. (1985) "Bootstrap Prediction Intervals for Regression", *JASA* **80**, 1026-31.

VTT (Technical Research Centre of Finland) (2000) *Nordic Treasure Hunt: Extracting Energy from Forest Residues* (ed. Eija Alakangas), Otamedia Oy, Espoo.

APPENDIX I FOREST CHIP ENERGY PRODUCTION IN FOREST REGIONS IN YEARS 2003 – 2008 (year 2003 index)



APPENDIX II. The OLS –estimation results for inverted OLS –forecasts

	Model $R_{it} = \alpha_1 + \beta_1 H_{it} + \varepsilon_{1,it}$				Model $S_{it} = \alpha_2 + \beta_2 H_{it} + \varepsilon_{2,it}$			
Region	α_1	t-value	β_1	t-value	α_2	t-value	β_2	t-value
1	-1.70	-1.97	0.83	81.76	2.47	4.17	0.16	19.72
2	-2.54	-5.74	0.97	91.98	1.94	8.25	0.07	7.40
3	-2.54	-2.93	0.88	36.06	2.34	5.12	0.17	9.92
4	-3.91	-2.10	0.94	31.30	5.39	2.80	0.17	3.23
5	-3.03	-3.36	0.97	78.64	2.34	5.62	0.05	5.09
6	-1.16	-1.67	0.64	53.69	1.76	5.00	0.36	39.59
7	-2.63	-3.27	0.90	29.11	1.37	5.05	0.38	14.29
8	-2.15	-2.84	0.96	131.47	2.46	5.85	0.05	8.16
9	-2.07	-5.64	1.01	56.06	0.14	0.57	0.70	16.98
10	-6.44	-6.37	0.94	32.75	5.05	5.59	0.21	5.83
11	-4.67	-2.57	0.96	30.06	3.56	4.43	0.11	3.95
12	-3.94	-4.53	0.79	32.72	3.11	8.72	0.23	14.57
13	-0.61	-0.62	0.53	4.75	-0.17	-0.82	0.94	39.22

APPENDIX III.

Fixed effects (FE) model estimates for trend model

$$H_{g,t} = a_g + \beta t + \varepsilon_{g,t}$$

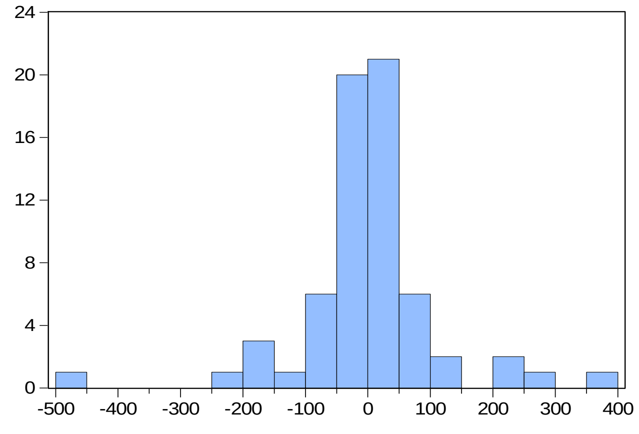
Dependent Variable: H_g (FOREST CHIPS)
Method: Panel Least Squares
Sample: 2003 - 2007, Cross-sections = 13, T = 5
Total panel (balanced) observations: 65
White cross-section standard errors & covariance (d.f. corrected)

	Coefficient	Std. Error	t-Statistic	Prob.
YEAR	54.762	23.13	2.36	0.022

Cross-section fixed (dummy variables)

REGION	FIXED EFFECTS	T-VALUES
1	592.57	7.05
2	364.53	4.34
3	283.85	3.38
4	377.91	4.49
5	204.17	2.43
6	385.87	4.59
7	115.93	1.38
8	670.19	7.97
9	40.91	0.49
10	246.18	2.93
11	94.65	1.13
12	123.73	1.47
13	8.82	0.10

R-squared	0.778	Mean dependent var	379.47
Adjusted R-squared	0.722	S.D. dependent var	240.77
F-statistic	13.807	Durbin-Watson stat	1.75



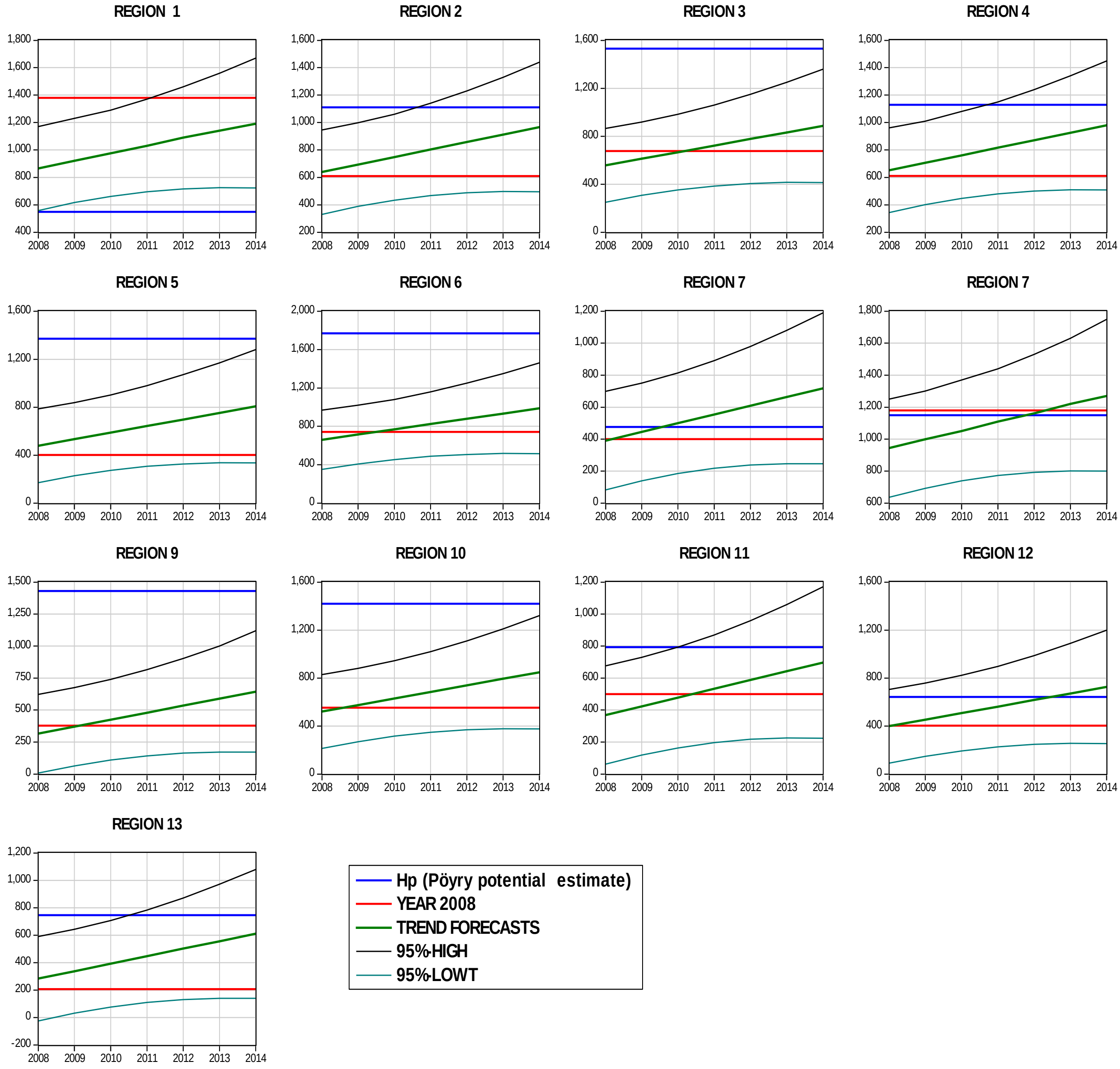
Series: Standardized Residuals
Sample 2003 2007
Observations 65

Mean	1.75e-15
Median	8.108377
Maximum	366.4910
Minimum	-476.2330
Std. Dev.	113.2548
Skewness	-0.496844
Kurtosis	8.387339
Jarque-Bera	81.27934
Probability	0.000000

APPENDIX IV

2008 – 2014 forest chips output forecasts with 90% CI's based on A) estimated forest region fixed effects, and B) on year 2008 observations as fixed effects.

A) FORECASTS WITH ESTIMATED FIXED EFFECTS



B) FORECASTS WITH YEAR 2008 FIXED EFFETS