BI Norwegian Business School - campus Oslo

GRA 19502

Master Thesis

Component of continuous assessment: Thesis Master of Science

Size management at regulatory thresholds by Norwegian companies

Navn:	Just ChristianHeide, Knut Østtveit Aardal

Start: 02.03.2017 09.00

Finish: 01.09.2017 12.00

Authors: **Just Christian Heide**
Knut Østtveit Aardal

Master Thesis

BI Norwegian Business School

- Size management at regulatory thresholds by Norwegian companies -

Supervisor:
John Christian Langli

Hand-in date:
31.08.2017

Programme:
Master of Science in Business
Major in Business Law, Tax and Accounting

Table of Contents

Table of figures, tables and formulas:

Executive summary

Discrete regulatory thresholds can incentivize companies to adjust size. In this thesis, we examine size management by Norwegian companies around thresholds for audit exemption, size classification and VAT-reporting. We adapt recently developed methodology on bunching. Our sample consist of data from all reported financial statements of Norwegian limited liability companies in the period of 1994 to 2015.

We identify significant size management around the revenue threshold for audit exemption, but it is not found around the other thresholds. In the period 2012-2015, there are approximately 14% more company-years than expected just below the revenue threshold, but the economic effect is limited. The average adjustment is estimated to approximately 147,000 NOK, adding up to a total of 58 mNOK. We are not able to identify one dominant method of size management, but our analysis indicate that companies reduce input rather than underreport revenues. The behavior is temporary and does not appear to significantly limit growth. Furthermore, we find a higher proportion of size management in unaudited companies and cash-intensive industries.

1 Introduction

In order to improve the business environment for small and medium companies, governments often simplify regulations for companies below certain size thresholds, such as a set level of revenue, total assets or employees.

Common examples of threshold-based regulations are VAT-registration, audit and disclosure requirements. As these requirements often incur considerable costs, companies could have an incentive to manage size below, or in some cases above, the thresholds. The exact level of the threshold usually has some element of arbitrariness, as the underlying measurements are continuous without a natural level that mandates the threshold. If size management is not considered, the threshold might be set at a suboptimal level.

To our knowledge research related to size management around thresholds in Norway is limited. However, several recent European studies indicate that size management is relatively common. Bernard, Burgstahler and Kaya (2014) found a higher than expected proportion of companies below thresholds for audit and disclosure requirements in several European countries. Almunia and Lopez-Rodriguez (In press) found similar behavior in Spain in order to avoid stricter tax enforcement. In a working paper, Harju, Matikka and Rauhanen (2016) uncover that companies manage size below the VAT-registration threshold in Finland, and that it may hinder growth among small companies.

In this thesis, we adopt recently developed methodology in order to examine the extent of size management in Norway and identify characteristics of companies that are likely to engage in the behavior. We focus our analysis on size management around the audit exemption thresholds: revenue, total assets and number of employees. In addition, we run tests on thresholds related to reduced government reporting and disclosure requirements; and VAT. Our research question is:

> *"Do Norwegian limited liability companies manage size to avoid statutory audit and other governmental regulations?*

1

2 Background

2.1 Audit exemption

In 2011, the government introduced voluntary audit exemption for small companies fulfilling certain criteria. Limited liability companies with revenue below 5 mNOK, total assets below 20 mNOK and a maximum average number of employees of 10 can choose to be exempted from audit requirements the following year (Asl, 1997, §7-6). A motivating factor for the relaxed regulation was reducing administrative burdens for small companies. Several European countries have had similar laws for some time, although the threshold varies considerably. Article 34 of the European Union directive 2013/34/EU dictates an upper limit of 6 mEUR in balance sheet total, 12 mEUR in net turnover and 50 employees (Federation of European Accountants, 2016). There appears to be no consensus on the optimum level, and thresholds in the Nordic countries are among the lowest.

The implementation of the voluntary audit has been evaluated by John C. Langli (2015). He concluded on overall that the negative consequences were limited, but noted that the market may not be in equilibrium yet. He highlights that the law affects a large number companies, but a low fraction of the economy is left unaudited. Further, he found that the quality of the tax returns among opt-out companies had declined significantly. However, most of this reduction was reversed for companies with an external accountant. We will examine if an external accountant can have a similar effect on the extent of size management.

In a recent effort to further decrease administrative burdens for small companies, the Norwegian Official Report (NOU) 22 from 2016 evaluates potential adjustments of the Limited Liability Companies Act (Aksjeloven) §7-6, including whether the audit thresholds should be increased. The committee did not recommend an increase, arguing that companies affected by a potential increase in the thresholds for revenue are largely companies in retail and construction. The Norwegian Tax Authority portray these industries as high risk when it comes to tax evasion. They are also overrepresented in the national bankruptcy register (Konkursregisteret) (NOU 2016:22, 2016, p. 162). Moreover, Norwegian companies in construction; accommodation and food service industries have been identified to have the lowest perception of the risk of discovery by the Tax Administration (Næringslivets

sikkerhetsråd, 2015). Following this argument, certain industries might be overrepresented in the group of companies below the audit thresholds.

2.2 Small companies

In Norway, small companies are subject to less regulation and have relaxed financial disclosure requirements. To be classified as a small, Norwegian companies must fulfil two of the three following requirements in two consecutive years; revenue below 70 mNOK, total assets below 35 mNOK and an average number of employees below 50 (Rskl, 1998, §1-6)[1]. Previous research on European companies suggests that companies manage size to avoid expanded disclosure of financial statements (Bernard et. al 2017). However, the studied European thresholds primarily concern disclosure of income statements, while Norwegian regulations concern disclosure of notes and cash flow statements.

2.3 VAT reporting

Size management around VAT-registration thresholds has been examined in Finland (Harju et al., 2016). Their results indicate that size management is caused by the compliance cost of VAT-reporting rather than the change in tax rate. In Norway, VAT-registered companies are required to file a VAT-return 6 times each year, but companies with revenue below 1,000,000 NOK in the past 12 months can apply for annual filing (Skatteforvaltningsforskriften, 2016, §8-3-3)[2]. Although there is no change in tax liability, the compliance costs may be reduced with less frequent filings. In addition to the administrative burden of filing the forms, the VAT-return requires that the accounts are reconciled and kept up-to-date. The threshold presents an opportunity to test whether the assumed change in compliance cost induces size management.

---

[1] The thresholds were increased July 1, 2005 and July 1, 2010. Consequently, we run tests on two periods. First 2011-2015 with the thresholds: revenue 70 mNOK, total assets 35 mNOK and an average number of employees of 50. Second 2006-2009 with the thresholds: revenue 60mNOK, total assets 30 mNOK and an average number of employees of 50.

[2] The threshold is regulated in *FOR-2001-06-29-800* in 2001-2009 and *Merverdiavgiftsloven §15-3* in 2009-2016.

3 Literature review

3.1 Earnings management

Research on size management around regulatory thresholds is an emerging field of study, but it builds on established literature on earnings management and accounting quality. Healy and Wahlen has often been cited for their definition of earnings management:

> *Earnings management occurs when managers use judgements in financial reporting and in structuring transactions to alter financial reports to either mislead some stakeholders about the underlying economic performance of the company or to influence contractual outcomes that depend on reported accounting numbers.* (Healy & Wahlen, 1999, p. 368)

Earnings management is a much-researched subject and is often associated with modification of accruals, changes in accounting methods and capital structure (Jones, 1991; Schipper, 1989). Healy and Wahlen emphasize that it is likely that earnings management is occurring, but the extent is difficult to determine. They identified three motivations for earnings management, which we will discuss below.

Previous research have suggested that influencing expectations of investors and capital markets can motivate managers to manipulate earnings. Trueman and Titman (1988) explained how managers are motivated to smooth income to cover the variance of the firm's underlying economic earnings. The authors theorize that this is because managers think that investors value firms with a smooth income stream higher. Burgstahler and Eames (2006) found evidence for upward earnings management to avoid reporting earnings below analysts' forecasts.

Different types of contracts present another motive for earnings management. Debt covenants and management compensation programs often rely on financial indicators that, to some degree, are determined at the management's discretion. A study by Sweeney (1994) documents a relationship between changing accounting practices as the firms are approaching the debt covenants constraints. Several studies have examined management's incentives to manage earnings as a response to executive bonus plans. Guidry, Leone, and Rock (1999), Healy (1985) and Holthausen, Larcker, and Sloan (1995) show how managers are likely to use

accounting judgement depending on earnings-based bonus schemes. As bonus schemes are often limited to an interval, the manager receives no additional bonus for increasing earnings above the maximum cap. Additionally, if the starting point for the bonus is not met, there is no punishment for further decreasing earnings. When earnings are outside the bonus interval, the manager can defer income or accelerate write-offs to maximize the bonus in later periods. Healy (1985) refer to manipulation above the maximum bonus target as "smoothing" income, and below the minimum bonus target as "taking a bath". This bears some similarity to our setting, where companies can be tempted to postpone revenue to stay below the threshold for consecutive years ("smoothing"), followed by an aggregated reversal of the deferred income when it is no longer possible to continue smoothing.

Other motivational factors are related to government regulations. Companies of a monopolistic nature (Cahan, 1992) or companies engaged in imports (Jones, 1991) could use income-reducing accruals to prevent exposure to unfavorable regulations. Other research examines companies adjusting to tax related regulations. Guenther (1994) found evidence of earnings management in response to the Tax Reform Act of 1986 in the United States. The reform reduced the maximum corporate tax level from 46% to 34 %. The author found a significant level of negative accruals in the year prior to the tax rate reduction. In other words, companies adjusted accruals to reduce the profit in the year with the highest tax rate, then increased profits when the tax rate was reduced.

Schipper (1989, p. 101) notes that research on earnings management must be interpreted with caution. Many empirical studies find indications of earnings management, but it is difficult to document the extent and exact methods of manipulation.

3.2 Discontinuities in size distributions

Several other studies have looked at inconsistencies in distributions of financial indicators. These studies are usually not targeted at uncovering the motives behind earnings management, but may be better suited to show the extent of management. Among firms in New Zealand, Charles Carslaw (1988) found that managers tend to round up earnings. He documented a higher than expected frequency of zeroes and

less than expected frequency of nines in reported positive earnings[3]. Jacob K. Thomas (1989) found similar tendencies in reported earnings and earnings per share among U.S. companies.

Burgstahler and Dichev (1997) examined the distribution of earnings in a large sample of American companies, and found that companies tend to manage earnings when pre-managed earnings are slightly negative. Their test assumes that the distributions are "smooth" in the absence of earnings management. The actual distributions are discontinuous around zero earnings, and the shape indicates that slightly negative earnings are managed to be positive. The authors also find a similar tendency in the distribution of year-over-year change in earnings, stating that the results "provide compelling empirical evidence that earnings decreases and losses are frequently managed away".

Durtschi & Easton, (2005, 2009) do not find a similar discontinuity around zero when examining undeflated earnings from the same data source. The authors claims that Burgstahler and Dichev's (1997) deflation of earnings distort the underlying distribution of net earnings and question the validity of the conclusion that discontinuity in the distribution of price-deflated earnings is evidence of earnings management. According to the authors, the discontinuity in the distribution is caused by "deflation, sample selection and a difference between the characteristics of profit and loss observations".

Other researchers have different explanations. Beaver, McNichols & Nelson (2007) claims that the discontinuity at zero earnings is caused by asymmetric effects of income tax and special items for profit versus loss companies. The article also corrects the calculation of variance in the Burgstahler-test (1997), claiming that the variance is understated and causes an overstatement of the standardized difference test statistic.

Burgstahler and Chuk (2015) later refuted Durtschi's (Durtschi & Easton, 2005, 2009) arguments and claim that Durtschi's research design shifts the focus to segments of the population where evidence of discontinuity is weak and inconclusive. They simulate the theoretical expectation of the standard difference of a normal

---

[3] In the second from left-most digit.

distribution and claim that Beaver et al.'s (2007) adjustment of their test does not improve, but rather result in an overstatement of the variance by about ⅓. Durtschi's critique against Burgstahler is mainly related to the scaling of earnings that the authors use to compare companies of different sizes. However, scaling is less relevant when comparing companies of similar size.

Burgstahler and Dichev's (1997) methodology has later been used to identify size management by Bernard, Burgstahler and Kaya (2017). They found evidence that private firms in several European countries manage size variables such as revenue, total assets and employees to stay below thresholds related to disclosure and audit requirements. As most of the countries have much higher thresholds than Norway, the mechanisms of management might differ. The compliance cost might be relatively higher for a small company with few employees than in a larger company with higher in-house competence. In addition, the audit thresholds in Europe tend to follow a 2:1 ratio between the revenue and asset threshold, while the Norwegian thresholds have a 1:4 ratio (Federation of European Accountants, 2016). While most European companies are limited by the asset threshold (Bernard et. al 2017), the revenue threshold is of higher practical importance in Norway.

3.3 Bunching

Emmanuel Saez (2010) introduced a new approach to threshold studies, which we will refer to as "bunching". The term relates to a higher than expected proportion of observations in the vicinity of a threshold. Saez analyzed the distribution of personal tax returns in the US and found bunching around the income level where federal tax liability starts. The extent of bunching is used to estimate the elasticity of reported income with respect to the marginal tax rate.

Chetty, Friedman, Olsen and Pistaferri (2011) use bunching to estimate the parameters in a labor supply model. They quantify the extent by estimating a counterfactual distribution that simulates the distribution without the presence of bunching. The bunching mass is defined as the excess mass (difference between actual and counterfactual distribution) around the kink point[4] between the actual and the counterfactual distribution.

---

[4] Kink point: the income level where the marginal tax rate increases

An uncertainty in Saez' and Chetty's studies is that optimization frictions, such as adjustment and attention costs, are not accounted for and could prevent individuals from bunching around thresholds (Henrik Jacobsen Kleven, 2016). Kleven and Waseem (2013) adjusts this by studying a setting where it is strictly better to keep income below thresholds. They use data on self-employed and wage-earners in Pakistan, where the tax is determined by a number of fixed average tax brackets. This system creates strong incentives for adjustments of income. As noted in the study: "at an income of 500,000 PKR, one more rupee of income triggers tax liability of 12,500 PKR for the self-employed". Consequently, there are intervals where taxpayers would be strictly better off by decreasing income. The authors argue that if these intervals are populated, it must be due to optimization frictions.
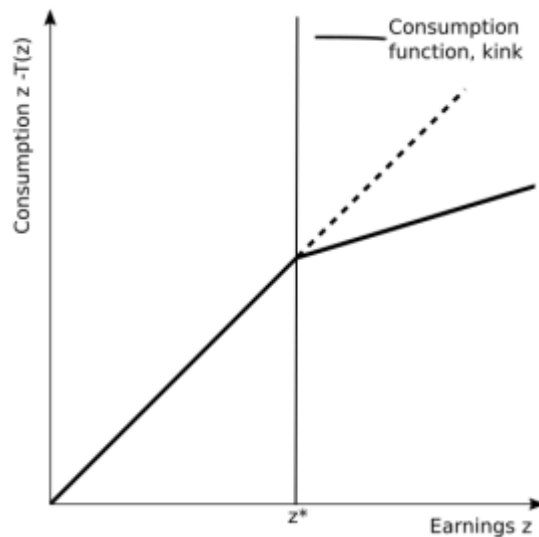
Figure 1: Consumption function, kink

Figure 2: Consumption function, notch

According to Kleven, the bunching approach "uses bunching around points that feature discontinuities in incentives to elicit behavioral responses and estimate structural parameters" (2016, p. 436). Kleven distinguishes between two different types of designs, "kinks" and "notches". The former relates to a "kink" in the consumption function, as only the earnings above the threshold z^* have an increased tax rate (Figure 1). On the other hand, a "notch" increases the tax liability on all earnings, leaving an interval z^* to z^{high} where individuals are either strictly better off or strongly incentivized to decrease earnings below the threshold (Figure 2).

8
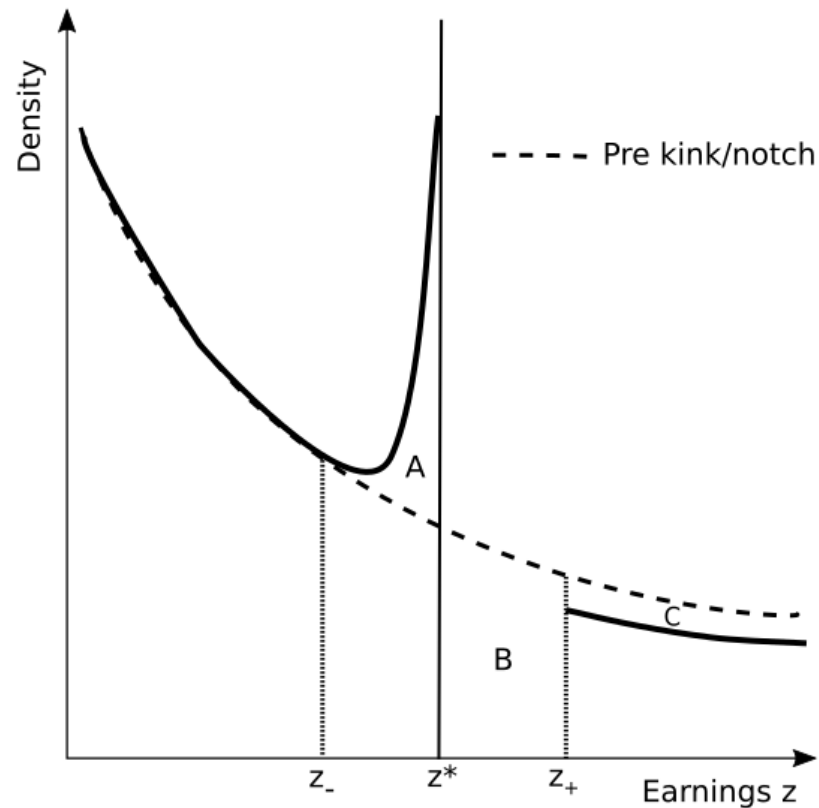
Figure 3: Density distribution

As earnings above the kink point are more aggressively taxed, fewer individuals will be inclined to earn more than the threshold z^*, which can be seen in the density distribution as bunching from area "B" to "A" (Figure 3). As the marginal tax rate are increased above the kink, individuals in area "C" will have incentives to reduce earnings and will fill up area "B". A theoretical post-kink distribution will according to Kleven have the form of Figure 4. In the case of a notch, individuals in "B" will have a strong incentive to bunch to area "A", but individuals in area "C" does not benefit from reducing their earnings to area "B". The average cost increases suddenly at z^*, not progressively above z^* as in a kink setting. Thus, the theoretical post-notch distribution will leave area "B" with no individuals, as illustrated in Figure 5. Empirical kink distributions have been shown to resemble the distribution in Figure 6, where the density is higher on both sides of z^*. This has been attributed to an inability to perfectly determine earnings (Chetty et al., 2011; Saez, 2010). The empirical notch distributions resemble Figure 7 due to optimization frictions and heterogeneous individual preferences (Henrik J Kleven & Waseem, 2013).
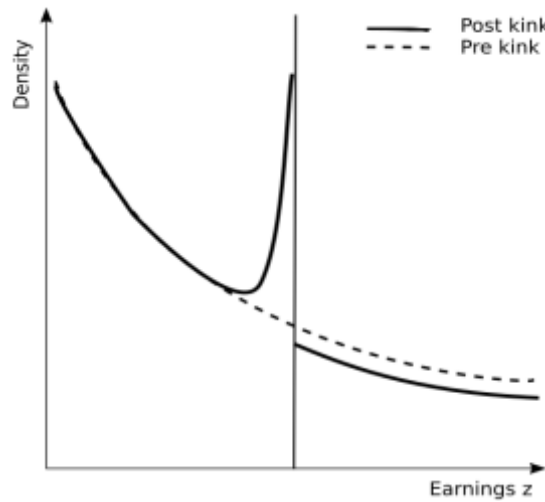
Figure 4: Density distribution without optimization friction, kink
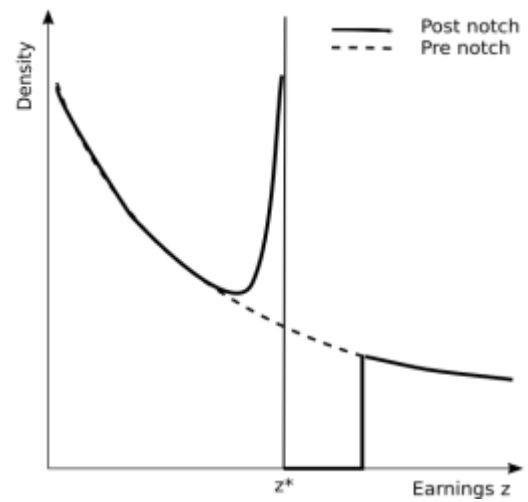
Figure 5: Density distribution without optimization friction, notch
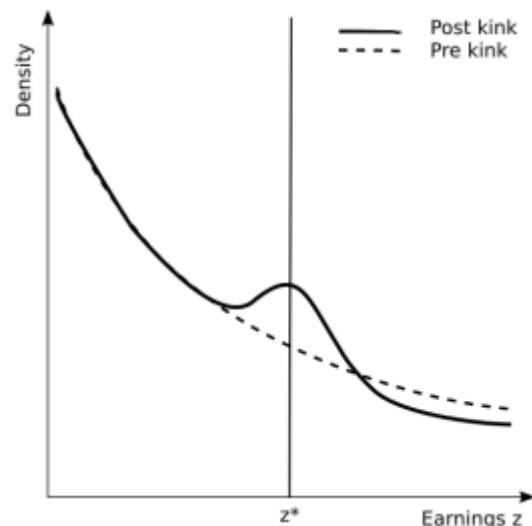
Figure 6: Density distribution with optimization friction, kink
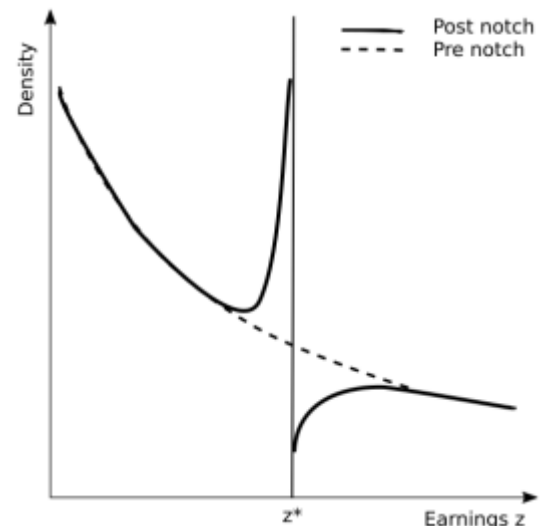
Figure 7: Density distribution with optimization friction, notch

Although Saez (2010) and Chetty et al. (2011) originally developed the method to estimate behavioral responses to taxes, the method has applications to other fields of research. Harju, Matikka and Rauhanen (2016) use bunching methodology to examine the sales threshold of 8,500 EUR for VAT-registration in Finland. In addition, they use variations in the VAT-rate and reporting requirements to distinguish the effects of compliance versus tax costs. The study uses data from the Finnish Tax Administration on all types of businesses in the period 2000-2013, combined with various other firm- and owner-level variables. The authors find sizeable bunching below the threshold, but the bunching behavior appears to be

caused by compliance costs rather than the increased tax liability. This is exemplified through two different changes in the VAT-system. First, a change in 2004 that decreased tax liability for firms just above the threshold, and second a simplification in the VAT-reporting in 2010. While the former did not significantly affect excess bunching, the latter decreased bunching by around 40%. Another interesting result is that the bunching behavior appears to be very permanent and hinders growth. Over 20% of firms with sales just below the threshold remain in the same interval the year after, compared to around 10% for neighboring intervals. After 4 years, this persistence rate is almost 10% for the just-below interval, over twice the rate of neighboring intervals. In a previous version of the article (Harju, Matikka, & Rauhanen, 2015), the authors identify firm characteristics with an OLS regression. They find similar evidence of high bunching persistence rates, but only minor effects of industry and owner-level characteristics.

Almunia and Lopez examine firm behaviour as a consequence of the establishment of the "Large Taxpayers Unit" in the Spanish tax administration that monitors firms with more than EUR 6 million in revenues. They find clear bunching under this threshold, and the response is stronger in sectors where transactions leave more paper trail. In addition, firms monitored by the LTU report larger tax bases. The authors conclude that the LTU is effective at reducing tax evasion, and that the optimal threshold would be lower than the current one. (Almunia & Lopez-Rodriguez, In press).

To our knowledge, the only research on bunching behavior around the audit threshold in Norway is a Master's thesis by Larsen and Løchen (2015). They examined the revenue distribution of limited liability companies in the period 2006 to 2013. They found signs of bunching starting in 2011, the same year as voluntary audit was implemented. The authors use methodology from Chetty et al. (2011) to quantify the extent, but the applied calculation method is unsuitable for the setting. In Larsen and Løchen's setting, the bunching companies are assumed to adjust from just above to just below the threshold, leaving a missing mass above the threshold (notch setting). They use Tore Olsen's (Chetty, 2011; 2012) Stata-program *"bunch_count"* which is written for a kink setting, where bunching is allowed to be present at both sides of the threshold. By using this program in a notch setting, the missing mass is

erroneously subtracted from the excess mass, which does not make sense theoretically, resulting in a deflated and invalid estimation of bunching. Hence, the authors have not been able to quantify the excess mass. In addition, their intervals are set at an uneven number of missing companies above the threshold compared to the excess mass below the threshold, which is not intuitive. If the masses had been equal on both sides of the threshold, the unfortunate application of the "*bunch_count*"-program would have revealed itself as the bunching would have been calculated to zero. Thus, a direct application of *"bunch_count"* is not appropriate in a notch setting.

Although Larsen and Løchen identify the bunching behavior and attempt to quantify the bunching, they do not intend to discover how companies bunch below the threshold. However, they used an OLS-regression to describe potential characteristics of companies just below the threshold. The characteristics differ marginally compared to the control groups, and they conclude that companies in the bunching regions are likely to have been in the same region in one of the two previous years. On an industry-level, they find that companies in health and social work are less likely to bunch.

3.4 Growth rate

Langli (2016) examines whether opt-out companies have lower revenue growth or lower operating profits than audited companies, which may indicate increased tax evasion. He finds that opt-out companies perform at least as good as the audited companies. In the period of 2011-2014 the revenue growth was significantly higher among the opt-out companies in three of the four years.

4. Methodology

4.1 Standardized difference tests

Discontinuities around thresholds for audit requirements has been documented by Bernard et al. (2017) in several European countries. We follow their methodology for distributional tests, which assume a smooth density distribution of the relevant variable in the absence of size management. The distribution is presented as a histogram, with the range of values divided in bins of equal width (Figure 8). The test is sensitive to bin width and how observations are distributed around the threshold

before manipulation. A narrow bin width improves accuracy, but the number of observations in each bin must be high enough to maintain the assumption of smoothness. We determine the bin width by visual inspection of the histograms and examine the sensitivity of the results with other choices of bin widths.
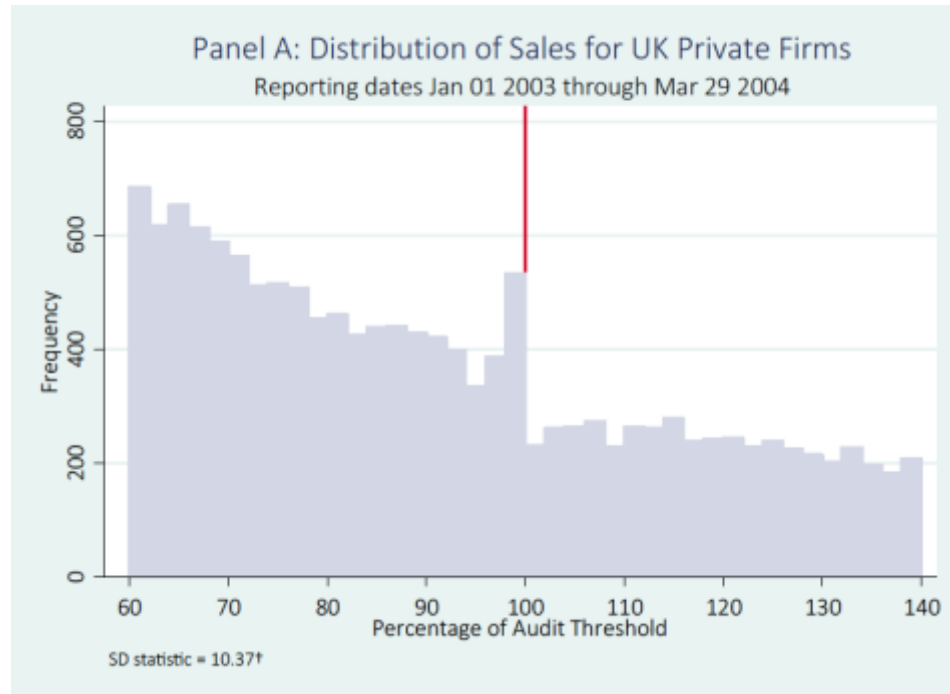
Figure 8: Distribution of sales for UK private firms around audit threshold.

Source: (D. Bernard et al., 2014)

Discontinuities around the thresholds are assessed using the previously discussed. The standardized difference-test for discontinuities were first defined in Burgstahler and Dichev (1997) and modified in Burgstahler and Chuck (2014). As the modifications have a limited practical impact, we will use the original test from 1997 (Formula 1):

$$SD = \frac{n_i - \frac{n_{i-1} + n_{i+1}}{2}}{\sqrt{3/2\, n_i}}$$

Formula 1: Standardized difference test statistic. Originated from (Burgstahler & Dichev, 1997), further explained in (Burgstahler & Chuk, 2014). We make the assumption that the variance expression assumes a smooth distribution.

where n_i is the number of companies in the bin just below the threshold. n_{i-1} and n_{i+1} are the number of companies in the bin just below and above n_i. In the case of a smooth distribution the average of n_{i-1} and n_{i+1} will approximate n_i with a small

13

deviation due to the convex form of the distribution. We account for the bin width sensitivity by expanding the test intervals, where n_i is tested against n_{i-2} and n_{i+2}; and n_{i-3} and n_{i+3}. Regarding the smoothness assumption, Burgstahler suggests that a visual inspection of the distribution is sufficient. We additionally run the SD-test for each bin the distribution (see example in Appendix C.1). If companies manage size to stay below the threshold, the standardized difference is expected to be positive.

4.2 Quantification of bunching

In order to quantify the bunching behavior of companies, we follow the methodology from Kleven (2016) and Harju (2016), which draws from the previous literature by Saez (2010), Chetty et al. (2011) and Kleven and Wasim (2013). The audit cost represents a discrete cost increase at the opt-out threshold, similar to the notch setting (Figure 9). We estimate the counterfactual distribution of companies by regressing a seven-degree polynomial function (Formula 2) to the observed distribution, and exclude the interval z_- to z_+ surrounding the threshold z^*.
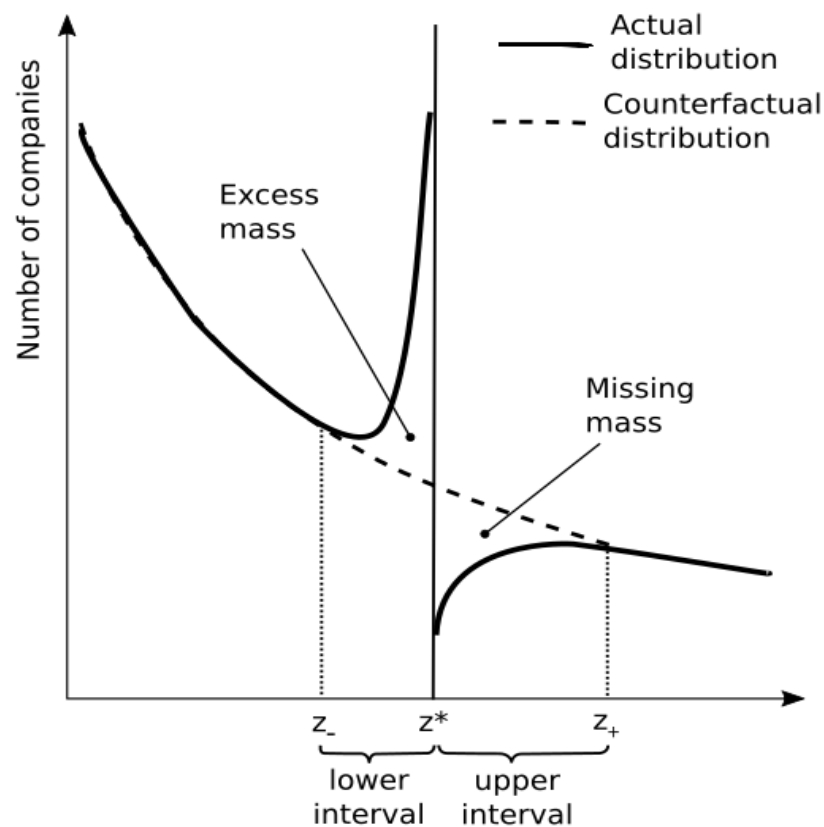
Figure 9: Bunching in a notch setting. Notation as in (Kleven, 2016).

$$c_j = \sum_{i=0}^{p} \beta_i \cdot (z_j)^i + \sum_{i=z_-}^{z_+} \gamma_i \cdot \mathbf{1}\left[z_j = i\right] + v_j,$$

Formula 2: Estimation of coefficients of the counterfactual distribution. Source: (Kleven, 2016)

where c_j is the number of companies in bin j, p is the order of the polynomial function and $[z_-, z_+]$ is the excluded interval. z_j is the level of the studied variable in bin j, e.g. revenue or total assets. β_j and γ_i denotes the coefficients, while v_j is the error term. By excluding the contribution from the dummies in the excluded interval, we predict the counterfactual number of companies in bin j, $\hat{c}_j$ (Formula 3).

$$\hat{c}_j = \sum_{i=0}^{p} \hat{\beta}_i \cdot (z_j)^i$$

Formula 3: Prediction of the counterfactual distribution. Source: (Kleven, 2016)

Bin width is determined by the same methodology as in the standardized difference test. By visual inspection, we determine the lower bound z_- of the bunching interval as the first bin where the actual distribution starts to deviate from the counterfactual distribution. The bunching from above to below the threshold creates a missing and excess mass in the distribution, respectively. The upper bound z_+ can be more difficult to determine visually, as the missing mass tend to gradually fade towards the counterfactual distribution with no clear ending point. Kleven and Waseem (2013) deal with this issue by identifying z_+ as the point where the excess mass is equal to the missing mass. The counterfactual function is fitted through an iterative process adjusting the upper bound to ensure the excess mass equals the missing mass.

Bernard et al. (2017) use the same methodology as Kleven and Waseem (2013) to construct a counterfactual distribution. Further on, they use this counterfactual to estimate the average number of bins managed, by assigning the missing mass in the bins above the threshold to the bins below the threshold based on their proportion of excess observations. We will use the same methodology to roughly estimate the average managed amount.

5 Data

5.1 Data source

Norwegian limited liability companies are required to publish an annual report, comprised of an income statement, balance sheet, accompanying notes, and a Board of Director's report. Although the annual report is publicly available, large-scale data collection are facilitated by third-parties. Our data is sourced from Norwegian Business School's Center for Corporate Governance Research (CCGR), which maintain an extensive collection of financial information on Norwegian companies. Our sample contains variables from all reported financial statements from Norwegian limited liability companies in the period 1994 to 2015. It contains 3,695,582 company-years from 448,287 companies. Additional variables in our dataset, related to companies that cannot be exempted from audit, are provided from the Supervisory Council for Legal Practice (Tilsynsrådet for Advokatvirksomhet) and the Financial Supervisory Authority of Norway (Finanstilsynet). A variable of whether companies use an external accountant are provided by The Norwegian Tax Administration (Skatteetaten).

5.2 Special concerns at the audit thresholds

There are some groups of companies that are not allowed to opt-out regardless of size, and therefore not incentivized to adapt to audit thresholds. For these thresholds, we exclude the following companies. 1) Pharmacies and law firms. The pharmacies are excluded based on the industrial codes provided by SSB as part of the CCGR-database and law firms according to the Supervisory Council for Legal Practice's register. 2) Approved entrepreneurs under the Lottery Act, according to reported NACE codes. 3) Companies under supervision of the Financial Supervisory Authority of Norway. 4) Parent and subsidiary[5] companies. Parent companies can not opt-out from audit. In addition, we exclude all subsidiaries as they might be less likely to opt-out as their parents already are subject to audit. Groups can also more easily adjust income and expenses between entities.

---

[5] It is unclear whether subsidiaries controlled by Norwegian Registered Foreign Companies (NUFs) are registered as subsidiaries by CCGR. The parent company is not mandated audit if total revenue is below 5 MOK. However, the number of companies affected by this is likely to be small.

The CCGR database contains several variables measuring average number of employees collected from different sources. Due to inconsistencies in how companies report this information and the extent of missing values we have not excluded companies based on this variable.

5.3 Descriptive statistics

The dataset contains a large number of companies for each year in the entire period. The observations per year are evenly distributed, with a small increase each year, ranging from 98,914 companies in 1994 to 262,975 companies in 2015 (Appendix A).

Whether a company is currently audited can affect the ability to adapt to thresholds. Consequently, we run separate tests for audited and un-audited companies. We exclude companies by the criteria described in 5.2. We define the post-audit-exemption period to be from 2012 - 2015 as the audit exemption regulation was introduced in May 2011. Companies only had 8 months to register, and one might only see a small effect of the new regulation during the first year. In 2015, about 60% of the financial statements were unaudited. Among companies with total revenue below 5 mNOK and total assets below 20 mNOK, the share was about 73%.

yr	audited 0	1	Total
2012	54,756	79,186	133,942
2013	72,182	70,283	142,465
2014	85,176	68,576	153,752
2015	97,960	66,596	164,556
Total	310,074	284,641	594,715

Table 1: Companies without (0) and with (1) an auditor 2011-2015 after exclusion of companies as described in section 5.2.

6 Analysis

6.1 Visual inspection

Before starting with statistical tests, we perform a visual inspection of the thresholds of interest. This gives us a rough view of the extent of bunching, whether the smoothness assumption holds, and indicates which bin width to use. The shape of the distributions helps us interpret and rule out spurious results from the SD-test.

6.1.1 Audit exemption - total revenue

The main assumption in Burgstahler's standardized difference(SD)-tests is that the pre-managed distribution is smooth. Figure 10 display the revenue distribution of companies in the period of 1994 to 2015 before excluding companies that are mandated audit. No clear visual signs of deviation from a smooth distribution can be seen.
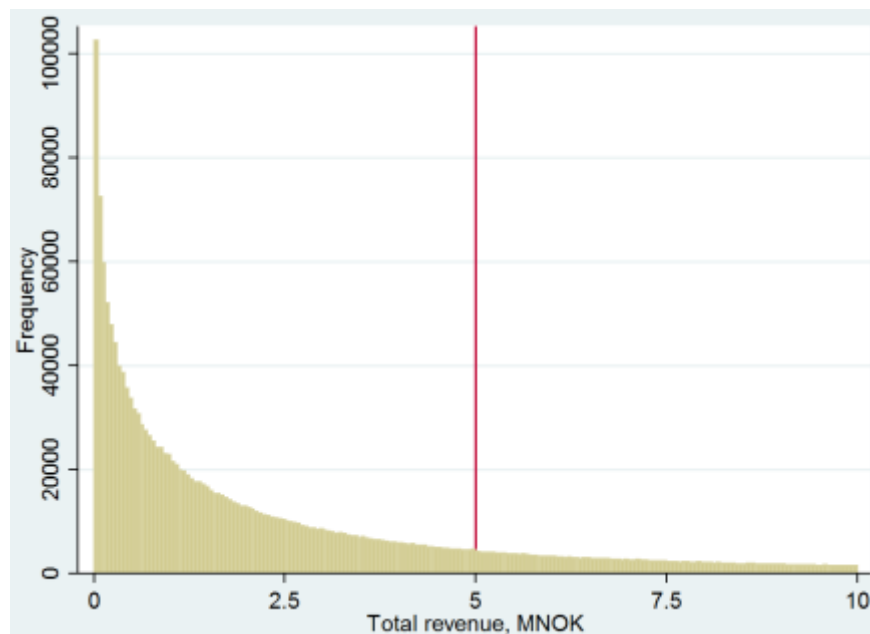
*Figure 10: Revenue distribution of companies in the period of **1994 - 2015**. Companies with revenue below 1,000 NOK and above 10 mNOK are excluded for scaling purposes.*

Figure 11a and b displays the revenue distribution before and after the audit exemption rules were implemented. Companies that are not allowed to opt-out are excluded (see section 5.2) in both periods. The shorter post-audit-exemption period

results in less observations, which again leads to a less smooth distribution[6].
Following a visual inspection and testing various bin widths as in Bernard et al.
(2017), we find that a width of 50,000 NOK is the best compromise between
accuracy and smoothness. As expected, there are no signs of discontinuity in the pre-
audit-exemption period. Afterwards, there is a sharp spike around the threshold of 5
mNOK, which indicates that companies are adapting to the threshold. The
discontinuity has gradually increased since 2011 (Figure 12).
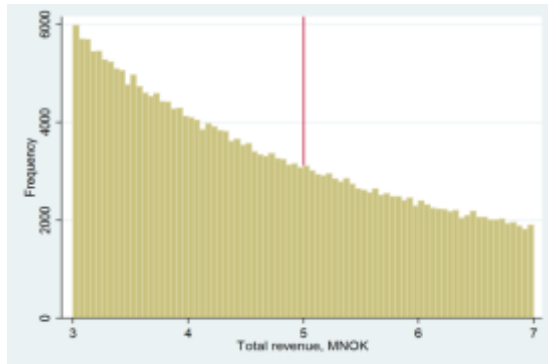
*Figure 11a: Revenue distribution of companies around audit threshold in the pre-audit-exemption period **1994 - 2010**.*
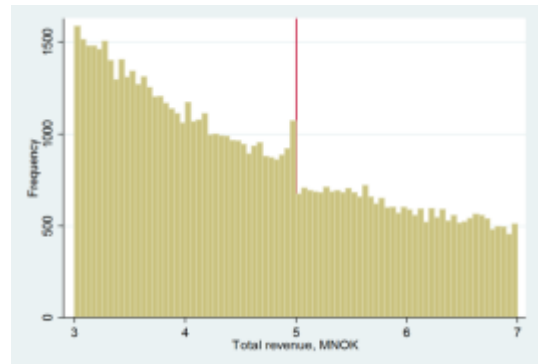
*Figure 11b: Revenue distribution of companies around audit threshold in the post-audit-exemption period **2012 - 2015**.*
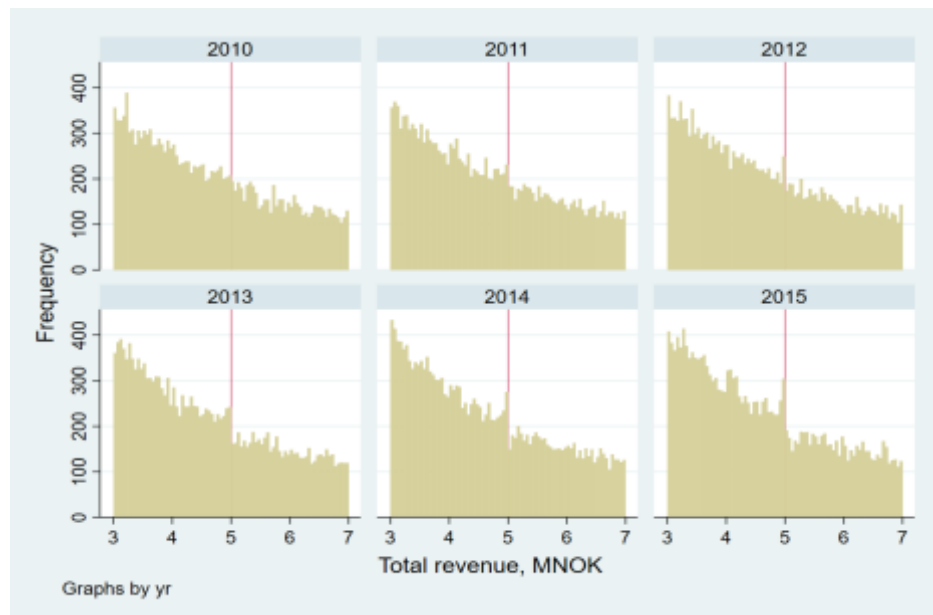
*Figure 12: Revenue distribution of companies around audit threshold **by year** in the period 2010-2015.*

---

[6] We have conducted the same test with an equal number of years on both sides of the threshold in the two periods with similar results.

This bunching tendency becomes even clearer when separating unaudited companies (Figure 13a) from audited companies (Figure 13b). Bunching tendencies only appear in the unaudited group. Audited companies that bunch below the threshold can opt-out from audit the same year and will thus not appear in Figure 13b, but in Figure 13a. The absence of bunching in Figure 13b indicates that audit is the only factor behind bunching at this threshold. The shape of the two distributions is affected by the number of firm years on each side of the threshold. Companies above the threshold in Figure 13a solely consist of growth companies (with revenue below 5 mNOK the previous year) which automatically are subject to audit in the subsequent year. The more horizontal distribution below the threshold in Figure 13b is caused by companies opting out from audit and correspondingly increasing the number of companies in the same interval in Figure 13a.
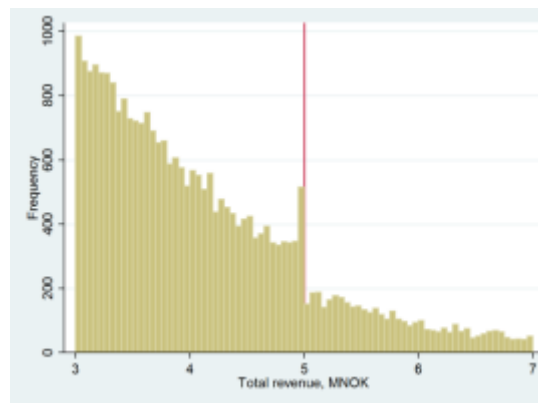
*Figure 13a: Revenue distribution of companies **without** auditor in the period of 2012 - 2015. Companies with total assets above 20 mNOK are excluded. Bin width 50,000 NOK.*
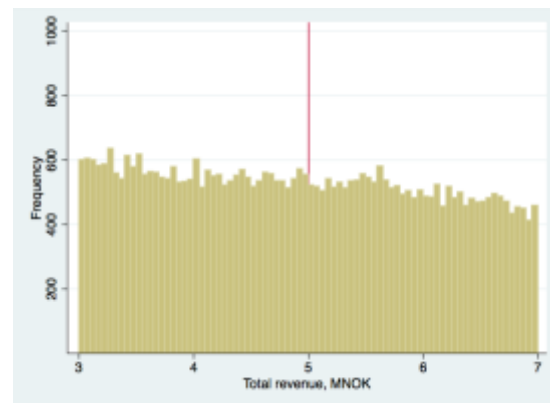
*Figure 13b: Revenue distribution of companies **with** auditor in the period of 2012 - 2015. Companies with total assets above 20 mNOK are excluded. Bin width 50,000 NOK.*

Norwegian companies have the choice of keeping accounting functions in-house or engage an external accountant. While there are no regulatory requirements on in-house personnel, the external accountant must be licensed, which requires a bachelors' degree and two years of relevant experience. This implies a higher level of competence, but a theorized effect on bunching is unclear. It might be that the increased competence is used to adopt to thresholds. Another possibility is that the external accountant prevents questionable accruals and inaccurate revenue recognition.

Figure 14a and b indicates that bunching appears to be present regardless of whether a company has an external accountant or not. However, the comparison

between the two distributions is uncertain because of the small number of companies without both auditor and external accountant[7].
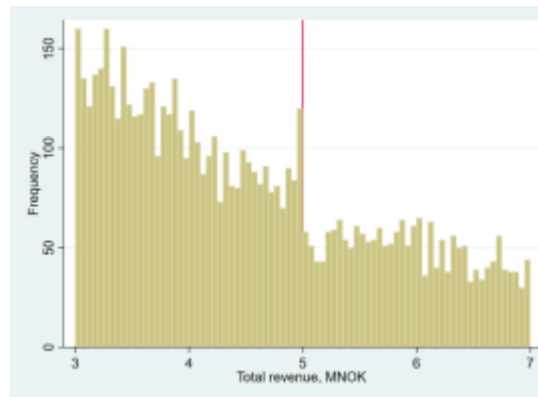
*Figure 14a: Revenue distribution of companies in the post-audit-exemption period 2012-2015, **without auditor and without external accountant**. Companies with total assets above 20 mNOK are excluded. Bin width 50,000 NOK.*
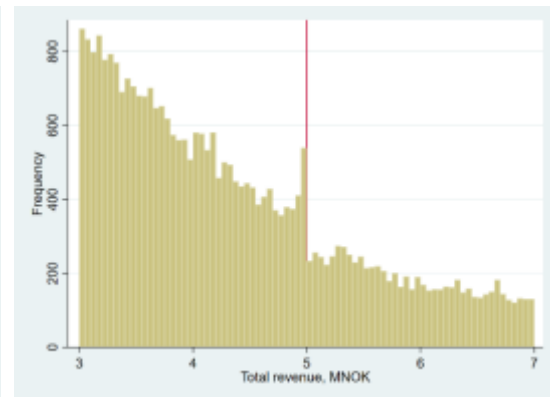
*Figure 14b: Revenue distribution of companies in the post-audit-exemption period 2012-2015, **without auditor and with external accountant**. Companies with total assets above 20 mNOK are excluded. Bin width 50,000 NOK.*

6.1.2 Audit exemption - total assets

A visual inspection of the total assets distribution of companies reveals no signs of bunching at the audit threshold of 20 mNOK, neither in the period of 1994-2010 (Figure 15a), nor in the period of 2012-2015 (Figure 15b). The small differences in frequency of the bins around the threshold are not remarkably different from the rest of the distribution. When we examine whether having an auditor affects the distribution, we do not visually detect any bunching (Appendix B.1 and B.2).

---

[7] Due to confidentiality requirements, we use an externally provided anonymized dataset for examining the effect of an external accountant. As we have not been in control of the data preparation, this sample is slightly different from our original sample.
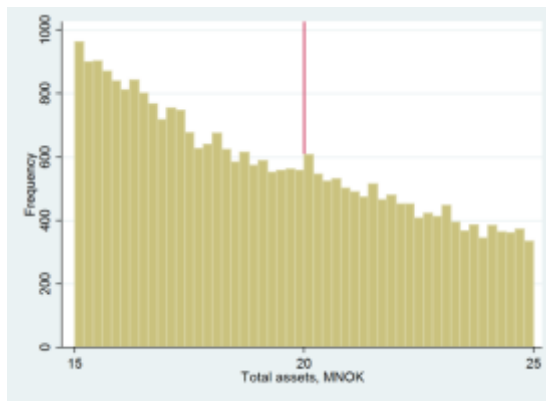
*Figure 15a: Distribution of companies by **total assets** in the pre-audit-exemption period of **1994-2010**. Companies with total revenue above 5 mNOK are excluded. Bin width 200,000 NOK.*
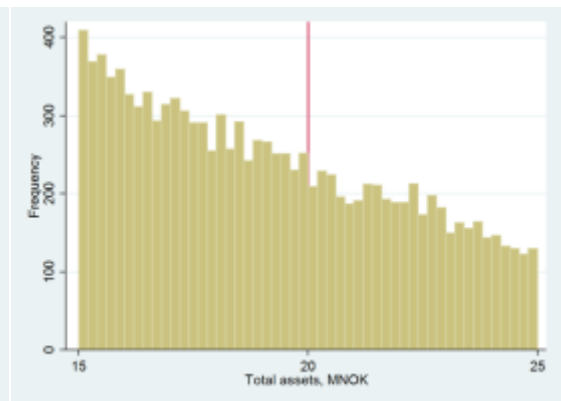
*Figure 15b: Distribution of companies by **total assets** in the post-audit-exemption period of **2012 - 2015**. Companies with total revenue above 5 mNOK are excluded. Bin width 200,000 NOK.*

6.1.3 Audit exemption - number of employees

We do not identify any bunching behavior in either periods when examining the average number of employees (Figure 16a and 16b). Whether the companies are audited or not does not seem to have any effect (Appendix B.3 and B.4). The high bin width relative to the threshold and the correspondingly high frequency of companies in each bin makes it difficult to observe discontinuities.
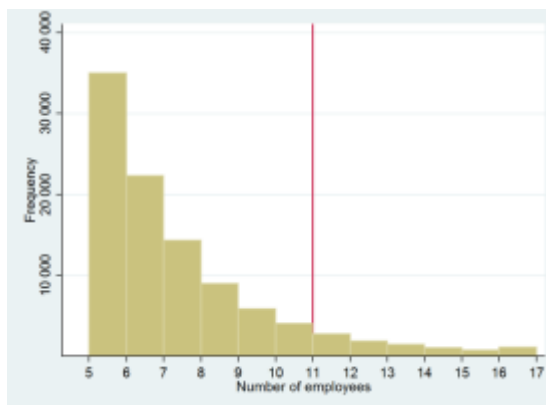
*Figure 16a: Distribution of companies by average number of employees in the pre-audit-exemption period of **1994-2010**. Companies with total assets above 20 mNOK and total revenue above 5 mNOK are excluded.*
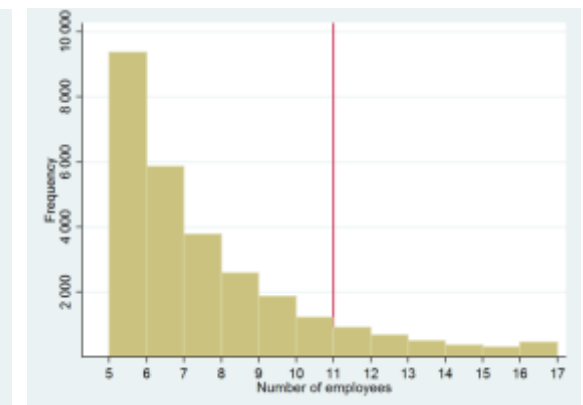
*Figure 16b: Distribution of companies by average number of employees in the post-audit-exemption period of **2012-2015**. Companies with total assets above 20 mNOK and total revenue above 5 mNOK are excluded.*

22

6.1.4 Small company classification

The revenue and assets thresholds for small company classification were adjusted in 2010. We find no signs of bunching around any of the thresholds, neither before (Appendix C.2) or after (Figure 17a, b and c) the change. This is in accordance with Bernard et al. (2017), who only found size management at thresholds that impose income statement disclosure. The Norwegian thresholds impose disclosure of a cash flow statement and expanded requirements, but income statements must be disclosed regardless of size.
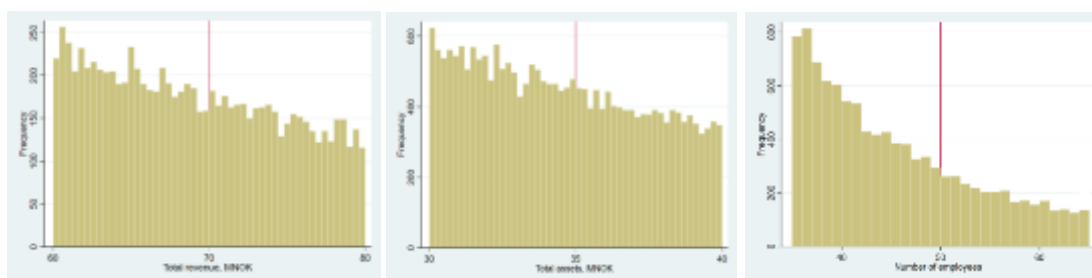
*Figure 17a: Total **revenue** distribution of companies in the period of 2011 - 2015 around the threshold for small companies 70 mNOK (if total assets < 35 mNOK or number of employees < 50). Bin width 400,000 NOK.*

*Figure 17b: Total **assets** distribution of companies in the period of 2011 - 2015 around the threshold for small companies 35 mNOK (if total revenue < 70 mNOK or number of employees < 50). Bin width 200,000 NOK.*

*Figure 17c: Average numbers of **employees** distribution of companies in the period of 2011 - 2015 around the threshold for small companies 50 average employees (if total revenue < 70 mNOK or total assets < 35 mNOK). Bin width 1.*

6.1.5 Annual VAT-reporting

Neither at the threshold for annual VAT-reporting does there appear to be bunching. The difference in compliance cost might be too small for companies to be willing to adapt to the threshold. Efficient digital solutions for VAT-reporting can explain the low compliance cost. The period examined covers years prior to the implementation of digital solutions, and we should be able to uncover discontinuities in these first years. However, an inspection of the revenue distributions year by year show no increased signs of discontinuities in the first periods which could have supported this hypothesis (C.3).
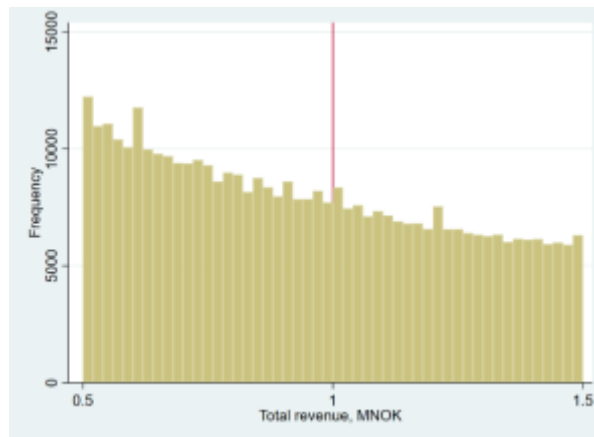
*Figure 18: Revenue distribution of companies in the period of **2002 - 2015** around the threshold for annual VAT-reporting 1,000,000 NOK. Bin width 20,000 NOK.*

6.2 Standardized difference test

After the visual inspection of the distributions, we perform SD-tests to examine the statistical significance of discontinuities at the thresholds. If companies are bunching, the standardized difference of the bin just below the threshold is expected to have a positive value, and a negative value in the bin just above the threshold. The critical values are 1.645, 2.236 and 3.090 for the significance levels of 0.05, 0.01 and 0.001, respectively (Suda & Shuto, 2007) .

The SD-test assumes linearity between the neighboring bins. Our distributions are convex, but the tested interval is approximately linear with a small enough bin size. This assumption is satisfied for the revenue and total assets distributions, but not for the average number of employees distributions related to audit threshold. A bin width of 1 does not create enough bins to satisfy the assumption of linearity between the neighboring bins. As a result, the standardized difference is systematically understated. Therefore, the results from test related to average number of employees around the audit threshold will be ignored.

6.2.1 Audit exemption thresholds

The results in Table 2 show no signs of bunching on a 0.05, 0.01 or 0.001 significance level at the audit thresholds in the pre-audit-exemption period. In the post-audit-exemption period, bunching at the revenue threshold is significant at a 0.001 level, with an SD of 6.83. The difference is even stronger for unaudited companies in the same period. For audited companies, none of the SD-values for

revenue are significant at a 0.05 level. These results imply that there is a correlation between bunching behavior related to revenue and the absence of an auditor. An intuitive explanation for this could be that the ability to adjust revenue is compromised by the auditor. A company that is aware of a forthcoming audit might be less willing to adjust accounting numbers. However, the characteristics of the groups differ. The audited companies around the threshold are generally older, have more assets and a lower growth rate, indicating that the groups are not directly comparable (Appendix G). It could be that the currently audited companies have a net positive gain from the audit, and thus less incentive to adjust revenue.

			Threshold	Bin width	No. of companies in bin -3 to +3	SD +/-1	SD +/-2	SD +/-3
Audit threshold 1994-2010	Total revenue	NOK	5 000 000	50 000	21 670	-0,80	0,06	-0,22
	Total assets	NOK	20 000 000	200 000	3 912	-0,97	0,21	0,66
Audit threshold 2012-2015	Total revenue	NOK	5 000 000	50 000	5 792	6,87***	6,87***	7,38***
	Total assets	NOK	20 000 000	200 000	1 646	1,67*	0,62	0,75
Audit threshold 2012-2015 with auditor	Total revenue	NOK	5 000 000	50 000	3 717	0,28	0,82	1,60
	Total assets	NOK	20 000 000	200 000	1 415	1,84*	0,61	0,99
Audit threshold 2012-2015 without auditor	Total revenue	NOK	5 000 000	50 000	2 075	9,60***	9,04***	8,99***
	Total assets	NOK	20 000 000	200 000	231	-0,14	0,14	-0,51
Audit threshold 2012-2015 without auditor with external accountant	Total revenue	NOK	5 000 000	50 000	2 435	7,65***	7,88***	8,00***
Audit threshold 2012-2015 without auditor without external accountant	Total revenue	NOK	5 000 000	50 000	516	3,65***	3,69***	4,73***
Simplified VAT reporting	Total revenue	NOK	1 000 000	20 000	55 031	-5,35	0,52	-0,10
Small companies 2006-2009	Total revenue	NOK	60 000 000	400 000	1 126	-0,53	-0,16	-0,89
	Total assets	NOK	30 000 000	200 000	2 475	-1,53	-0,86	0,13
Small companies 2011-2015	Total revenue	NOK	70 000 000	400 000	1 215	-0,71	-1,04	-1,55
	Total assets	NOK	35 000 000	200 000	3 140	0,93	1,12	1,79*

Table 2: Standardized difference-tests for the bin just below the threshold. SD +/-1, SD +/-2 and SD +/-3 refer to the standardized difference of a bin relative to the closest, second closest and third closest pair of neighboring bins. The statistically significance on 0.05, 0.01 and 0.001 levels are denoted with *, ** and ***, respectively. Similar statistics for bins above the thresholds are presented in Appendix E.1.

In the post-audit-exemption period, the SD-value for total assets is significant on a 0.05 level for the closest neighboring pair of bins, but not significant at any level for the other neighboring pairs of bins. The values are approximately the same for companies with an auditor, but not significant for unaudited companies. These results give no, or at best very weak, evidence of bunching. The SD-test gives only statistically significant results for one of the three neighboring pairs, which can indicate that the identified deviation is caused by a less smooth distribution rather than bunching behavior.

For the three audit exemption thresholds, we find significant bunching for revenue, weak bunching tendencies for total assets, and inconclusive results for number of employees. A potential explanation for this can be the low threshold for revenue relative to other thresholds. Among the companies with revenue below 5 mNOK, only 4% have total assets above 20 mNOK (and 2% have more than 10 employees on average). For most companies, revenue seems to be the only threshold of any practical importance.

6.2.2. Annual VAT-threshold and small company threshold

Regarding annual VAT-reporting, we find no signs of bunching under the threshold. There is a discontinuity just above the threshold, but the SD-statistic (7.00) is invalid due of a lack of smoothness. When calculating SD-values for all bins in the interval of 500,000 - 1,500,000 NOK we find numerous cases of significant SD values (Appendix C.1). A narrower bin width gives discontinuities at several points with a distance of 50,000, which is a typical round number effect (Appendix C.4). However, we also find discontinuities at other values which we are not able to explain. The visual inspection of the histogram suggest that the fluctuating SD-values can be caused by unexplained characteristics of the sample, rather than a weakness in the SD-test. However, the absence of bunching below the threshold suggest that the VAT-filing process is efficient enough to prevent compliance cost exceeding the cost of adjusting revenue.

Nor around the thresholds for small company classification do we find any statistically significant bunching. To be considered as a small company, only two of the requirements must be fulfilled and they have to be met for two consecutive years. These requirements make it harder to find significant results, because the potential

bunching behavior is diluted. In addition, the threshold for revenue are four times larger relative to total assets compared to the audit exemption threshold. A potential consequence is that companies to a larger extent needs to comply to all three requirements simultaneously, and thus makes it more difficult to bunch below the thresholds.

6.2.3 Findings

To conclude the visual inspection and SD-test, significant bunching is only found at the revenue threshold for audit exemption. In addition, the bunching tendencies is intensified when excluding companies that already have an auditor. As a result, we will solely examine the extent of bunching and identification of the bunching companies at the revenue threshold.

6.3 Bunching by industry

To further identify the bunching companies, we conduct the same SD-test combined with visual inspection, on industry-level. The companies are segmented into sections and divisions based on NACE-codes (Statistisk Sentralbyrå, 2008). Through the visual inspection, we find that there are too few observations to maintain a smooth distribution with a bin width of 50,000 NOK. Hence, we increase the bin width to 100,000 NOK, which lowers the accuracy but increases the smoothness of the distribution. The lowered accuracy impacts our ability to detect smaller adjustments of revenue. The sections Construction (F) and Wholesale and retail trade (G) emerge as the industries with the highest bunching tendencies, with statistical significant SD-values for all three neighboring pairs of bins on a 0.001 confidence level. Other notable industries are Transporting and storage (H) and Accommodation and food service activities (I) with SD-values significant on 0.01 and 0.05, respectively, for all three neighboring bins. Additionally, the sections related to service activities (M and N) show some signs of bunching.

NACE-Section	Section name	No. of companies in bin -3 to +3	SD +/-1	SD +/-2	SD +/-3
C	Manufacturing	705	0,56	0,36	0,28
F	Construction	2 282	5,53***	5,35***	4,65***
G	Wholesale and retail trade; repair of motor vehicles and motorcycles	3 319	4,35***	5,32***	4,49***
H	Transporting and storage	1 098	2,60**	2,61**	3,33***
I	Accommodation and food service activities	804	2,54**	2,16*	1,82*
J	Information and communication	486	0,88	1,12	0,47
L	Real estate activities	446	1,56	0,85	1,32
M	Professional, scientific and technical activities	1 137	2,70**	2,05*	0,77
N	Administrative and support service activities	544	2,14*	2,47**	1,24
Q	Human health and social work activities	850	-0,22	0,55	-0,18

*Table 3: Standardized difference-test of revenue **by NACE-section** for the bin just below the audit threshold in relation to its neighboring bins. SD +/-1, SD +/-2 and SD +/-3 refer to the standardized difference of a bin relative to the closest, second closest and third closest pair of neighboring bins. The statistically significance on 0.05, 0.01 and 0.001 levels are denoted with *, ** and ***, respectively. Bin width 100,000 NOK. Only sections with more than an average of 50 observations in the bins around the threshold are included.*

By breaking the NACE-sections F-I down into divisions we get a more detailed description of the bunching companies. Construction of buildings (41), Specialized construction activities (42), Motor vehicles (45) and Wholesale trade (46) stand out as the divisions with the highest presence of bunching. Retail trade (47) and Warehousing (52) has a smaller extent of bunching.

NACE-section	NACE-division	Division name	No. of companies in bin -3 to +3	SD +/-1	SD +/-2	SD +/-3
F	41	Construction of buildings	860	3,40***	2,63**	2,76**
F	43	Specialised construction activities	1 362	4,36***	4,81***	4,09***
G	45	Wholesale and retail trade and repair of motor vehicles and motorcycles	799	3,39***	4,31***	3,80***
G	46	Wholesale trade, except of motor vehicles and motorcycles	793	3,23***	3,29***	3,45***
G	47	Retail trade, except of motor vehicles and motorcycles	1 729	1,34	1,98*	1,01
H	49	Land transport and transport via pipelines	403	2,15*	2,10*	2,92**
H	52	Warehousing and support activities for transportation	383	-0,11	0,39	-0,17
I	56	Food and beverage service activities	679	2,35**	2,24**	1,42

*Table 4: Standardized difference test of revenue **by NACE-division** for the bin just below the audit threshold in relation to its neighboring bins. SD +/-1, SD +/-2 and SD +/-3 refer to the standardized difference of a bin relative to the closest, second closest and third closest pair of neighboring bins. The statistically significance on 0.05, 0.01 and 0.001 levels are denoted with *, ** and ***, respectively. Bin width 100,000 NOK. Only sections with more than an average of 50 observations in the bins around the threshold are included.*

6.3.1 Findings

Our results are consistent with a common perception that industries related to construction and other cash-intensive industries are more often affiliated with questionable business practices. However, we only found weak signs of bunching among companies in Retail trade (47). This is traditionally an industry with a high level of cash transactions. However, electronic payments have to a large extent replaced cash in recent years. Additionally, the retail sector has a high volume of transactions, which may be difficult to accrue.

6.4 Quantification of bunching
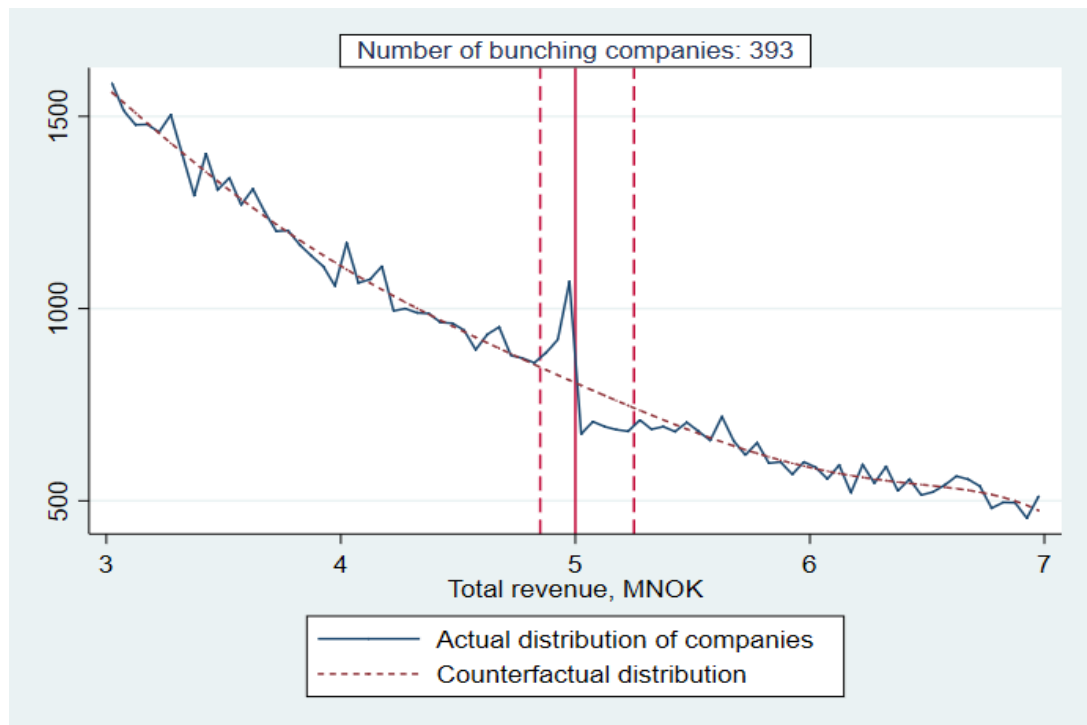
*Figure 19: Actual and counterfactual revenue distribution in the period **2012-2015**. The counterfactual is a seven-degree polynomial. Bin width is 50,000 NOK. See Appendix D for estimation results.*

Figure 19 displays the actual (solid) revenue distribution and the estimated counterfactual distribution (dashed) of all companies in the period of 2012-2015. The shape of the actual distribution is consistent with Kleven's (2016) notch setting. The lower limit of bunching is visually determined to start at 4,850,000 NOK, while the upper limit is set to 5,250,000 NOK, where the excess mass below the threshold equals the missing mass above. Consequently, the upper interval (250,000 NOK) is

wider than the lower interval (150,000 NOK) indicating that companies bunch from a larger section. Langli (2015) estimated the average net savings for opt-out companies in 2012 to be 19,250 NOK. This estimate was based on a different sample, with other firm years and exclusion criteria. Although not directly applicable, we assume an average net savings of 20,000 NOK for the companies in our sample. Combined with an average operating margin of 7,5%[8], companies would have to increase revenue by over 250,000 NOK to cover the cost of audit. However, only 10% of the companies in the upper interval bunch, meaning that 90% of the companies either value the audit or have adjustment costs higher than 19,250 NOK. The adjustments cost contains all costs, including non-monetary, related to adjustment. Companies that are unable to legally adjust revenue may be unwilling of moral reasons, while others can simply be unaware of the threshold.

We estimate an average bunching response of 147,527 NOK, by taking the sum of the weighted average number of bins managed multiplied with bin width (Appendix F). With 393 bunching company-years (14% more than expected from the counterfactual), the estimated total amount of adjusted revenue is approximately 58 mNOK. However, we are not able to determine if this revenue is lost or reversed in later periods.

7. Mechanisms of bunching

7.1 Evidence from firm-level factors

To understand the mechanisms of bunching, we examine how companies are adjusting to the threshold. We are not able to identify individual bunching companies, but we theorize that company characteristics in the bunching interval should differ from the neighboring intervals. However, the companies that are naturally located in this interval will dilute the difference and thus complicate the identification. Previous research attempting to identify characteristics of bunching companies has encountered challenges. Almunia and Lopez-Rodriguez (In press), Harju et al. (2016) and Larsen and Løchen (2015) attempt to identify characteristics using multiple regression models, but their results are ambiguous and the explanatory power is

---

[8] Companies in our sample with revenues between 4,5 and 5,5 mNOK.

weak. Instead, Almunia & Lopez-Rodriguez and Harju et al. focus on a more general bivariate analysis, which we follow in this section. The idea is to study changes in the relationship between revenues and various firm-level factors.

We follow Harju's method of creating bivariate graphs by running local polynomial smoothing on the selected variables on the y-axis over revenue on the x-axis. This smoothing technique is a non-parametric estimation which do not assume a fixed distribution of the underlying data. Stata suggest a rule-of-thumb bandwidth by default. We increase the bandwidth to 50,000 to increase the smoothing. We use Stata's default settings for kernel function (Epanechnikov) and local-mean smoothing (degree(0)). As the procedure is sensitive to outliers on the y-axis, we winsorize the variables on the y-axis at the upper and lower 1%.

Harju et al. (2016, p. 27) explains that "if evasion through underreporting of sales is the main explanation for how firms locate themselves below the threshold, we should find that the level of reported expenses, wages and equity levels are larger just below the threshold." However, the graphs in Figure 20 and 21 are approximately linear, with no significant discontinuities at the threshold. This suggests that underreporting of sales is not a dominant cause of bunching in our setting. A possible explanation for the linear relationship is that companies bunch by reducing output with a corresponding decrease in expenses, wages and equity.

When running the localized regressions on industry-basis, we find that sample sizes are too small to fit accurate estimations.

Expenses

Profits
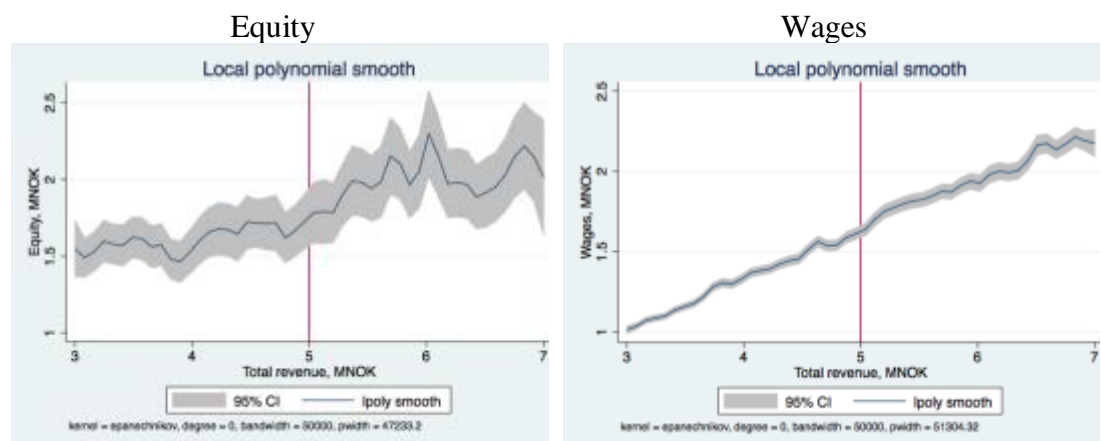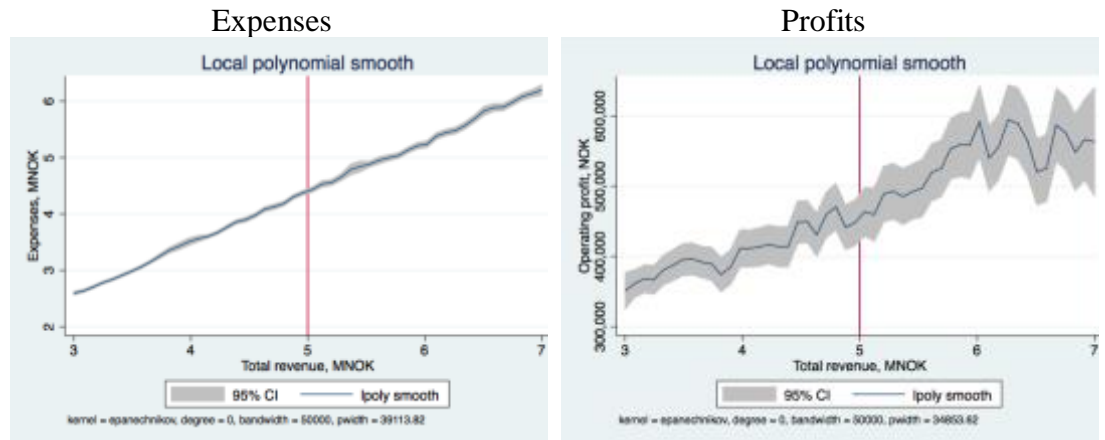
*Figure 20: Local regression on **firm-level factors** around the revenue threshold in the period 2012-2015. To account for outliers the variable on the y-axis is winsorized by setting the upper and lower 1% of observations equal the 1% and 99% percentile.*

Equity

Wages

Expenses

Profits
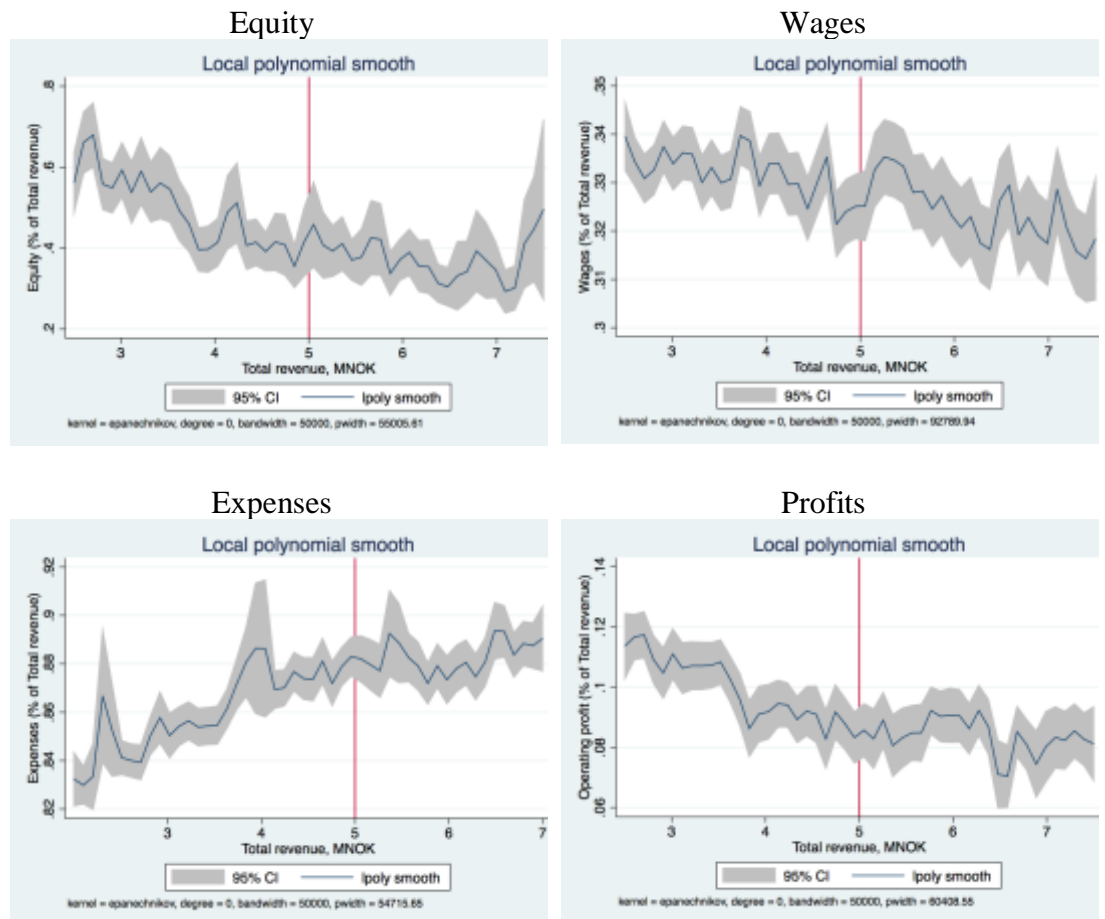
*Figure 21: Local regression on **firm-level factor rates** around the revenue threshold in the period 2012-2015. To account for outliers the variable on the y-axis is winsorized by setting the upper and lower 1% of observations equal the 1% and 99% percentile.*

7.2 Growth

Since companies bunch by adopting a revenue level that is lower than it would have been in a threshold-free environment, a key question is whether the thresholds hinder growth. If this is the case, we expect to find a lower growth rate among the companies just below the threshold. In addition, we should find a higher persistence rate[9] in the bin just below the threshold compared to the surrounding bins.

Figure 22 displays the logarithmic one-year growth in revenue in the period 2012-2015 plotted as a local polynomial function with a 95% confidence interval. Contrary to our expectations in a growth-hindering scenario, the growth rate drops at the threshold. A potential explanation is that the different frequencies of audited and un-audited companies in the bins around the threshold cause this change in growth rate. As stated in section 6.2.1, unaudited companies have higher average growth rates than audited companies. The higher frequency of un-audited companies below the threshold increases the average growth rate, while the lower frequency above the threshold lower the average growth rate.
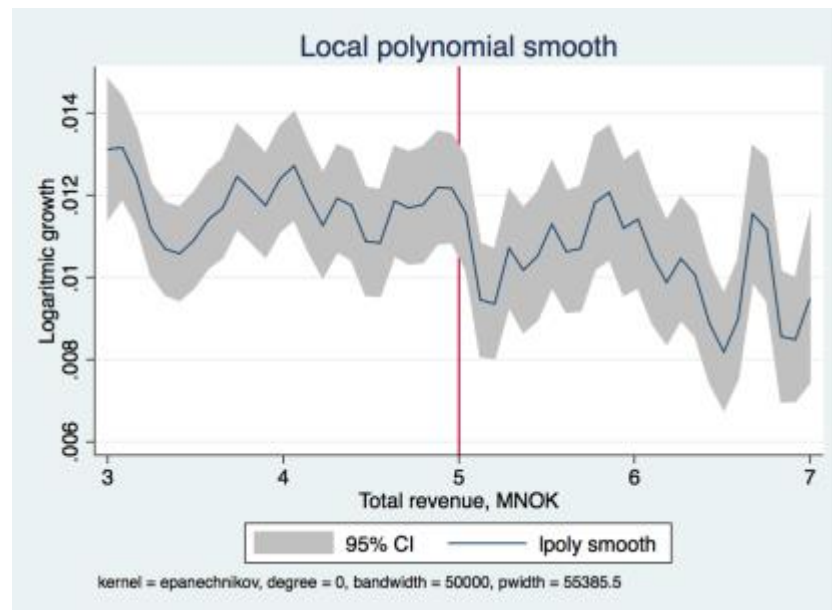
Figure 22: Logarithmic one year growth in revenue in the period 2012-2015. To account for outliers the variable on the y-axis is winsorized by setting the upper and lower 1% of observations equal the 1% and 99% percentile.

---

[9] Persistence rate: proportion of companies in a bin which was located in the same bin in previous years.

Figure 23 plots the proportion of companies located in the same bin as in a previous year. Although the one-year persistence rate just below the threshold is about 2 percentage points higher than neighboring bins, the gaps disappear in later years. This indicates that the individual bunching behavior is not permanent.
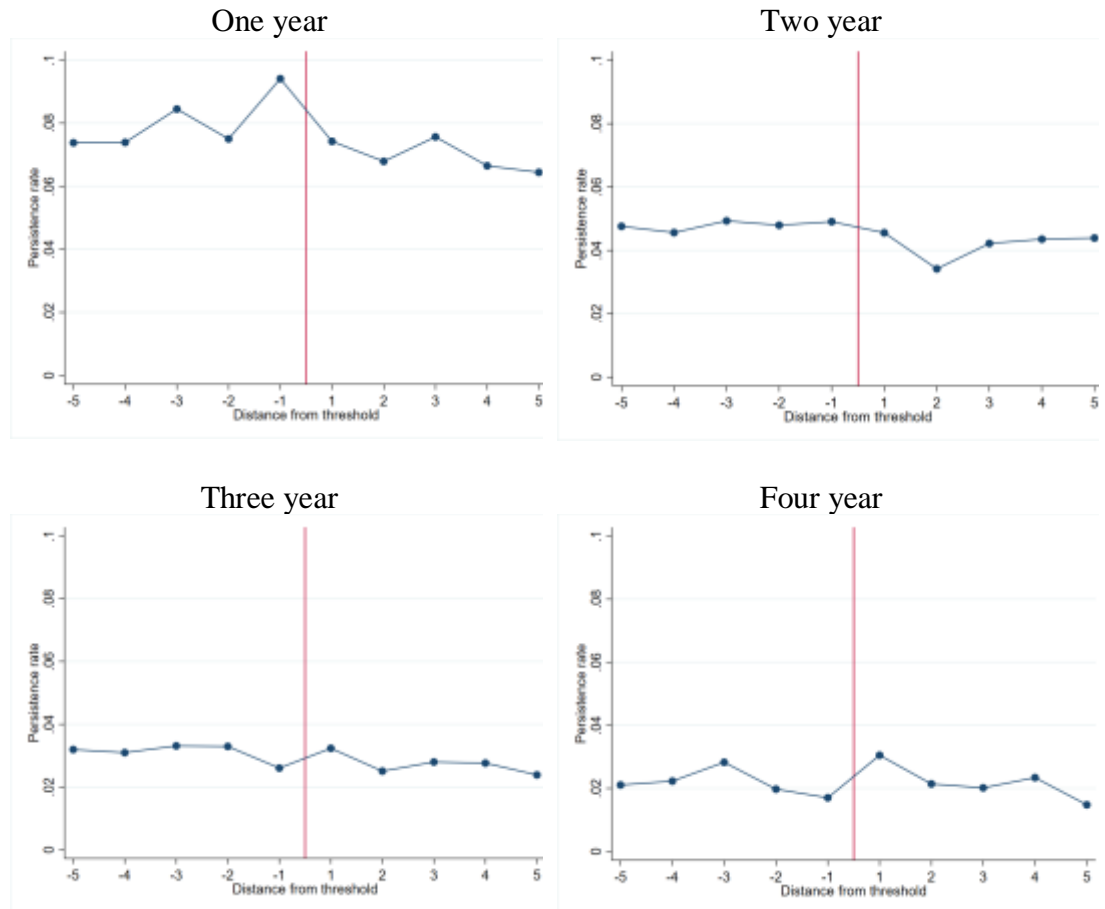
Figure 23: Persistence rate of companies being located in the same bin as one, two, three and four years ago. Bin width 150,000 NOK. Period 2011-2015.

As seen in Figure 24, the one-year persistence rate appears to have increased the later years. This may indicate that the effects of the audit exemption have not yet reached an equilibrium. It might be too early to conclude on the persistency of bunching.
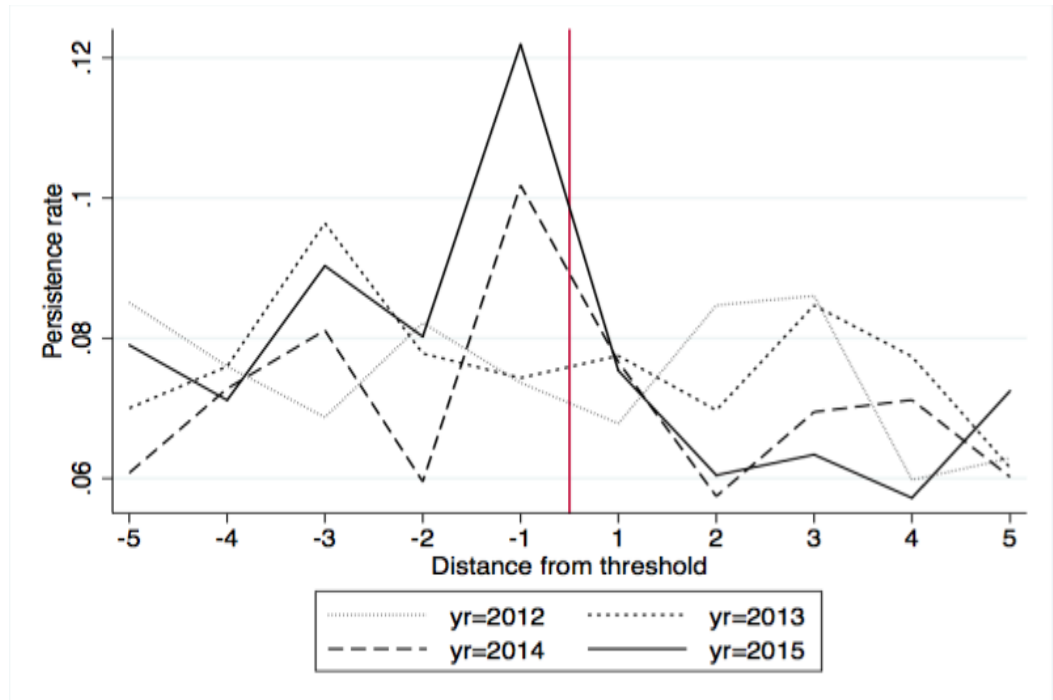
Figure 24: Persistence rate by year of companies located in the same bin as the previous year. Bin width 150,000 NOK. Period 2011-2015.

7.3 Additional remarks

Besides bunching, another possible method of not exceeding the audit threshold is to start new companies and continue operations in several separate entities. This response is difficult to observe in our dataset. With additional ownership-variables, one could examine whether the number of demergers or companies per owner have increased after 2011. However, the increased administrative costs affiliated with operating multiple companies is likely to exceed the cost savings of an audit exemption. This is especially the case with demergers, as an audit is explicitly required as part of the demerger process.

8. Conclusion

Previous studies have shown that bunching at thresholds is widespread in several European countries (Bernard et al. 2017; Harju et al. 2016). We quantify and provide evidence of bunching at the revenue threshold for audit exemption. Although the bunching is significant, the economic consequences are limited. We find approximately 14% more than expected company-year observations just below the threshold in the period 2012-2015. An estimated 58 mNOK in revenue were adjusted downwards, but we have not identified whether this is lost or reversed in later periods. Although our analysis does not provide clear evidence of the anatomy of bunching, it indicates that companies manage size by reducing output rather than underreporting revenue.

Bunching varies by industry. Especially cash-intensive industries are prominent, such as construction, repair and sale of motor vehicles, wholesale and transportation.

Companies do not bunch for an extended period of time. There is a slightly increased probability that a company just below the threshold will remain the next year, but the difference disappears the following years. However, our analysis indicate that this rate may be increasing. We have not found other signs of reduced growth among the bunching companies.

Bunching is not found among companies that are currently audited, indicating that the audit prevents bunching. However, we find it probable that this is rather caused by differences in characteristics. The currently audited companies are on average older, have lower growth and more assets. Motivations for choice of audit are likely different from currently un-audited companies, which affects the incentives for bunching. Whether a company use an external accountant does not appear to influence bunching.

We do not find evidence of bunching around other thresholds for audit exemption, annual VAT-reporting and size classification.

Literature

Almunia, M., & Lopez-Rodriguez, D. (In press). *Under the radar: the effects of monitoring firms on tax compliance*. Retrieved from http://www2.warwick.ac.uk/fac/soc/economics/staff/malmuniacandela/almunialopezmar2017-under_the_radar.pdf

Asl. (1997). Lov 13. juni 1997 nr 44 om aksjeselskaper.*, §7-6*.

Beaver, W. H., McNichols, M. F., & Nelson, K. K. (2007). An alternative interpretation of the discontinuity in earnings distributions. *Review of Accounting Studies, 12*(4), 525-556.

Bernard, D., Burgstahler, D., & Kaya, D. (2014). Size management by European private firms to minimize disclosure and audit costs.

Bernard, D., Burgstahler, D., & Kaya, D. (2017). Size Management by European Private Firms to Avoid Disclosure and Audit Costs.

Bernard, D. B., David; Kaya, Devrimi. (2017). Size Management by European Private Firms to Avoid Disclosure and Audit Costs. *Working Paper*.

Burgstahler, D., & Chuk, E. (2014). *Detecting earnings management using discontinuity evidence*. Retrieved from

Burgstahler, D., & Chuk, E. (2015). Do scaling and selection explain earnings discontinuities? *Journal of Accounting and Economics, 60*(1), 168-186.

Burgstahler, D., & Dichev, I. (1997). Earnings management to avoid earnings decreases and losses. *Journal of Accounting and Economics, 24*(1), 99-126.

Burgstahler, D., & Eames, M. (2006). Management of earnings and analysts' forecasts to achieve zero and small positive earnings surprises. *Journal of Business Finance & Accounting, 33*(5- 6), 633-652.

Cahan, S. F. (1992). The effect of antitrust investigations on discretionary accruals: A refined test of the political-cost hypothesis. *Accounting Review*, 77-95.

Carslaw, C. A. (1988). Anomalies in income numbers: Evidence of goal oriented behavior. *Accounting Review*, 321-327.

Chetty, R. (2011). Estimation of Bunching in Distributions.   Retrieved from http://www.rajchetty.com/utilities/

Chetty, R., Friedman, J. N., Olsen, T., & Pistaferri, L. (2011). Adjustment costs, firm responses, and micro vs. macro labor supply elasticities: Evidence from Danish tax records. *The quarterly journal of economics, 126*(2), 749-804.

Durtschi, C., & Easton, P. (2005). Earnings management? The shapes of the frequency distributions of earnings metrics are not evidence ipso facto. *Journal of Accounting Research, 43*(4), 557-592.

Durtschi, C., & Easton, P. (2009). Earnings management? Erroneous inferences based on earnings frequency distributions. *Journal of Accounting Research, 47*(5), 1249-1281.

Federation of European Accountants. (2016). *Audit exemption thresholds in Europe*. Retrieved from https://www.accountancyeurope.eu/wp-content/uploads/1605_Audit_exemption_thresholds_update.pdf

Guenther, D. A. (1994). Earnings management in response to corporate tax rate changes: Evidence from the 1986 Tax Reform Act. *Accounting Review*, 230-243.

Guidry, F., Leone, A. J., & Rock, S. (1999). Earnings-based bonus plans and earnings management by business-unit managers. *Journal of Accounting and Economics, 26*(1), 113-142.

Harju, J., Matikka, T., & Rauhanen, T. (2015). The Effect of VAT Threshold on the Behavior of Small Businesses: Evidence and Implications. *Helsinki, Finland: VATT Institute for Economic Research Working Paper*.

Harju, J., Matikka, T., & Rauhanen, T. (2016). The effects of size-based regulation on small firms: evidence from VAT threshold.

Healy, P. M. (1985). The effect of bonus schemes on accounting decisions. *Journal of Accounting and Economics, 7*(1-3), 85-107.

Healy, P. M., & Wahlen, J. M. (1999). A review of the earnings management literature and its implications for standard setting. *Accounting horizons, 13*(4), 365-383.

Holthausen, R. W., Larcker, D. F., & Sloan, R. G. (1995). Annual bonus schemes and the manipulation of earnings. *Journal of Accounting and Economics, 19*(1), 29-74.

Jones, J. J. (1991). Earnings management during import relief investigations. *Journal of Accounting Research*, 193-228.

Kleven, H. J. (2016). Bunching. *Annual Review of Economics, 8*, 435-464.

Kleven, H. J., & Waseem, M. (2013). Using notches to uncover optimization frictions and structural elasticities: Theory and evidence from Pakistan. *The quarterly journal of economics, 128*(2), 669-723.

Langli, J. C. (2015). *Evaluering av unntak for revisjonsplikt i små aksjeselskaper*. Retrieved from https://www.regjeringen.no/contentassets/981a92dc8e474d6ea18e4be36d860255/revisjonsplikt.pdf

Langli, J. C. (2016). Resultatføring av inntekter og kostnader før og etter fravalg av revisor i små AS–Tyder utviklingen på økte skatteunndragelser? *Praktisk økonomi & finans, 32*(02), 200-214.

Larsen, C. K., & Løchen, J. A. (2015). The Effect of Introducing Voluntary Audit on Accounting Quality and Firm Behaviour.

NOU 2016:22. (2016). *Aksjelovgivning for økt verdiskapning*. Retrieved from https://www.regjeringen.no/contentassets/5b2effd420504a5b9b72721e8402941d/no/pdfs/nou201620160022000dddpdfs.pdf

Næringslivets sikkerhetsråd. (2015). *Kriminalitets- og sikkerhetsundersøkelsen i Norge*. Retrieved from https://www.nsr-org.no/getfile.php/Dokumenter/NSR%20publikasjoner/Krisino/krisino_2015_utskrift.pdf

Olsen, T. (2012). Estimation of Bunching in Distributions. rajchetty.com. Retrieved from http://www.rajchetty.com/utilities/

Rskl. (1998). Lov 17. juli 1998 nr 56 om årsregnskap m.v. (regnskapsloven).

Saez, E. (2010). Do taxpayers bunch at kink points? *American Economic Journal: Economic Policy, 2*(3), 180-212.

Schipper, K. (1989). Commentary on earnings management. *Accounting horizons, 3*(4), 91-102.

Skatteforvaltningsforskriften. (2016). Forskrift 6. Februar 2017 nr 1360 til skatteforvaltningsloven.

Statistisk Sentralbyrå. (2008). Standard Industrial Classification.

Suda, K., & Shuto, A. (2007). Earnings management to meet earnings benchmarks: Evidence from Japan. In M. H. Neelan (Ed.), *Focus on Finance and Accounting Research* (pp. 67-85): Nova Science Publishers Inc.

Sweeney, A. P. (1994). Debt-covenant violations and managers' accounting responses. *Journal of Accounting and Economics, 17*(3), 281-308.

Thomas, J. K. (1989). Unusual patterns in reported earnings. *Accounting Review*, 773-787.

Trueman, B., & Titman, S. (1988). An explanation for accounting income smoothing. *Journal of Accounting Research, 26*, 127-139.

APPENDICES:

Appendix A: Number of Norwegian limited liability companies per year.

yr	Freq.	Percent	Cum.
1994	98,914	2.68	2.68
1995	105,026	2.84	5.52
1996	110,694	3.00	8.51
1997	118,360	3.20	11.72
1998	126,336	3.42	15.14
1999	130,301	3.53	18.66
2000	135,562	3.67	22.33
2001	138,176	3.74	26.07
2002	140,625	3.81	29.87
2003	141,502	3.83	33.70
2004	143,966	3.90	37.60
2005	157,264	4.26	41.85
2006	180,259	4.88	46.73
2007	191,627	5.19	51.92
2008	197,648	5.35	57.26
2009	199,889	5.41	62.67
2010	203,020	5.49	68.17
2011	208,049	5.63	73.80
2012	221,629	6.00	79.79
2013	235,546	6.37	86.17
2014	248,214	6.72	92.88
2015	262,975	7.12	100.00
Total	3,695,582	100.00	

Appendix B: Audit-threshold

B.1 Distribution of companies by total assets in the post-audit-exemption period of 2012 – 2015, without auditor. Companies with total revenue above 5 mNOK are excluded. Bin width 200,000 NOK.
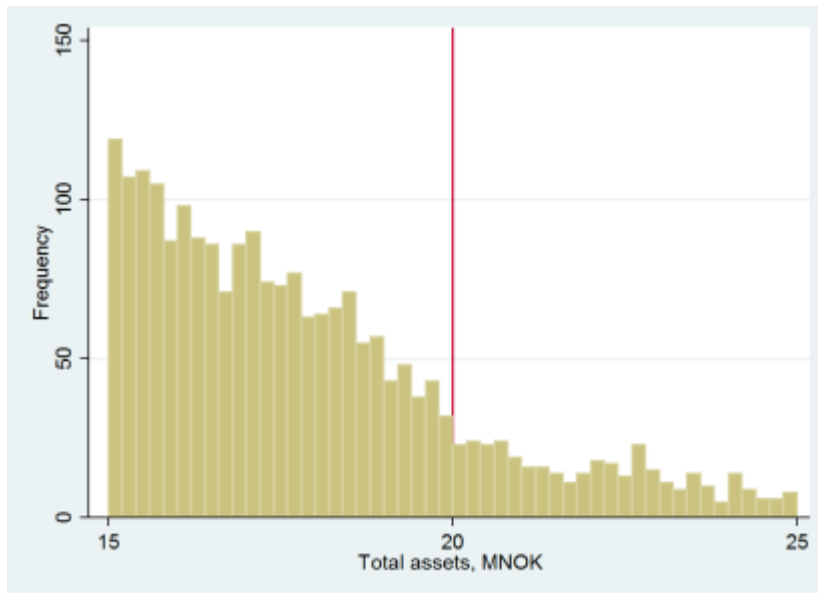
B.2 Distribution of companies by total assets in the post-audit-exemption period of 2012 – 2015, with auditor. Companies with total revenue above 5 mNOK are excluded. Bin width 200,000 NOK
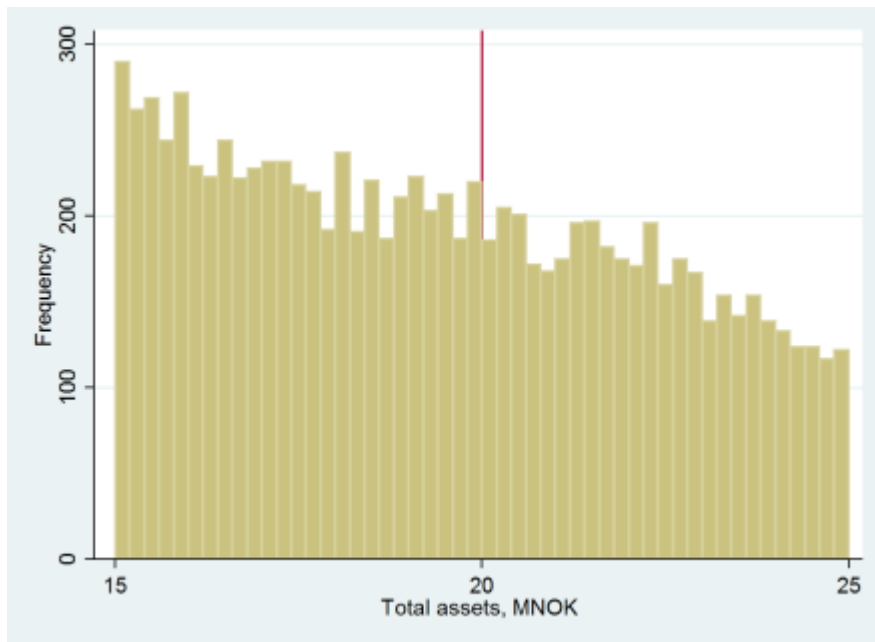
B.3 Distribution of companies by average number of employees in the post-audit-exemption period of 2012-2015, without auditor. Companies with total assets above 20 mNOK and total revenue above 5 mNOK are excluded.
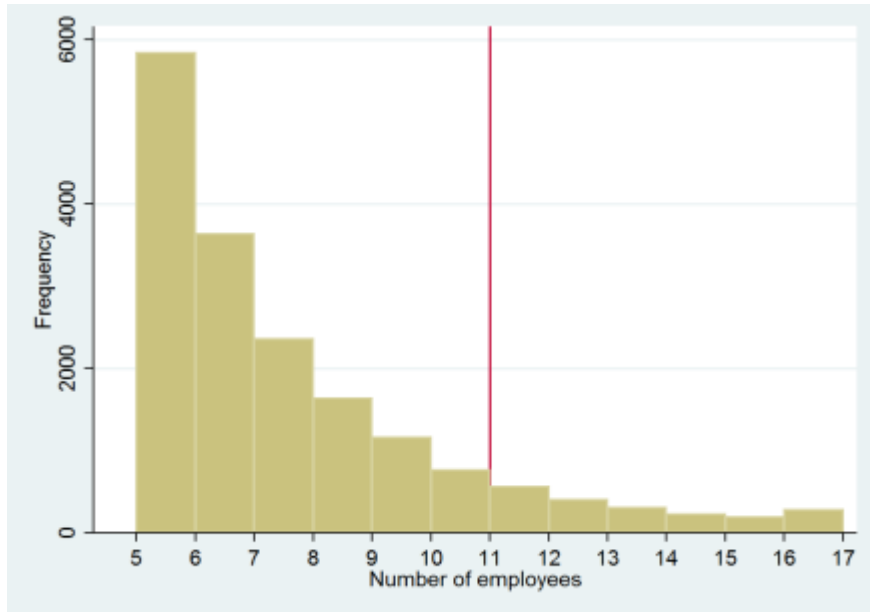
B.4 Distribution of companies by average number of employees in the post-audit-exemption period of 2012-2015, with auditor. Companies with total assets above 20 mNOK and total revenue above 5 mNOK are excluded.
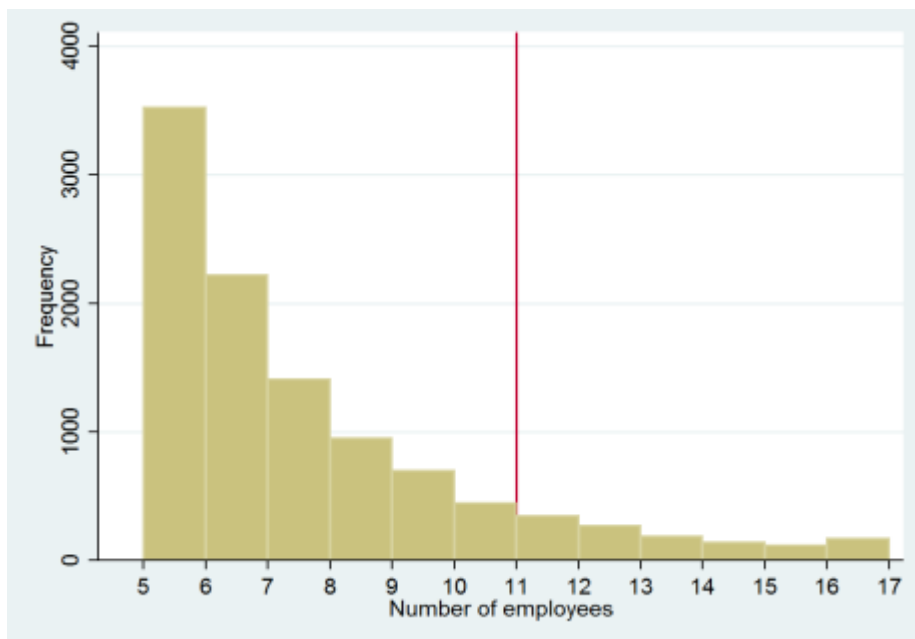
Appendix C: Annual VAT-reporting and small company classification

C.1

Standardized difference per bin, threshold for annual VAT-reporting (figure 18)
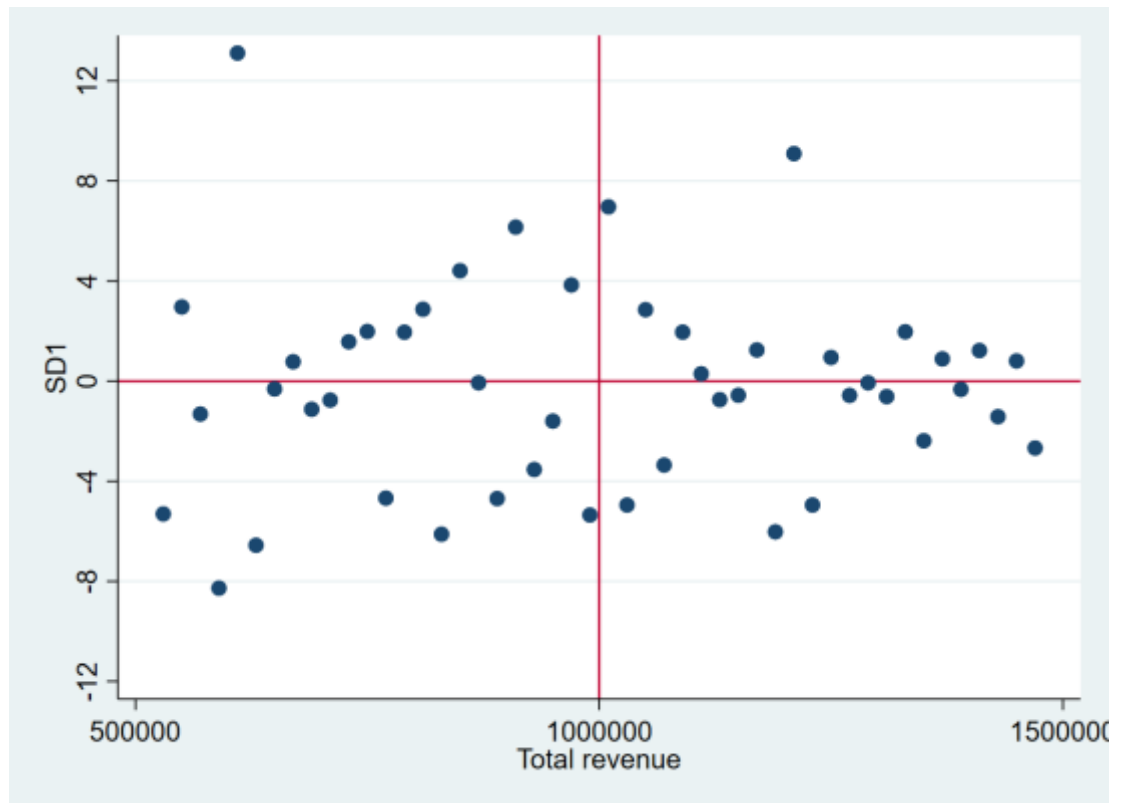
C.2
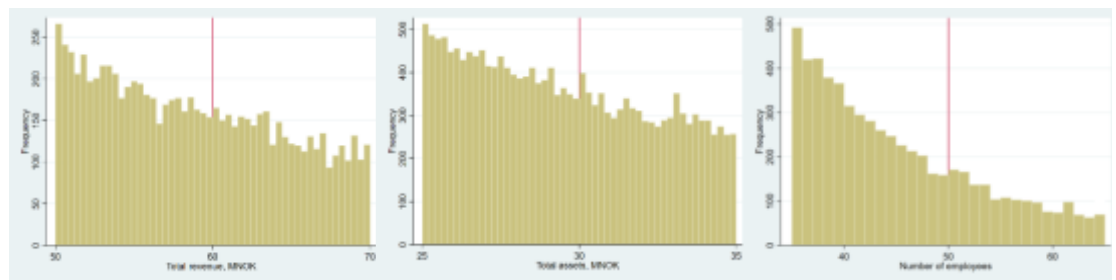
*Total **revenue** distribution of companies in the period of 2006 - 2009 around the threshold for small companies 60 mNOK (if total assets < 30 mNOK or number of employees < 50). Bin width 400,000 NOK.*	*Total **assets** distribution of companies in the period of 2006 - 2009 around the threshold for small companies 30 mNOK (if total revenue < 60 mNOK or number of employees < 50). Bin width 200,000 NOK.*	*Average numbers of **employees** distribution of companies in the period of 2006 - 2009 around the threshold for small companies 50 average employees (if total revenue < 60 mNOK or total assets < 30 mNOK). Bin width 1.*

C.3 Annual VAT-reporting, by year

Graphs by yr
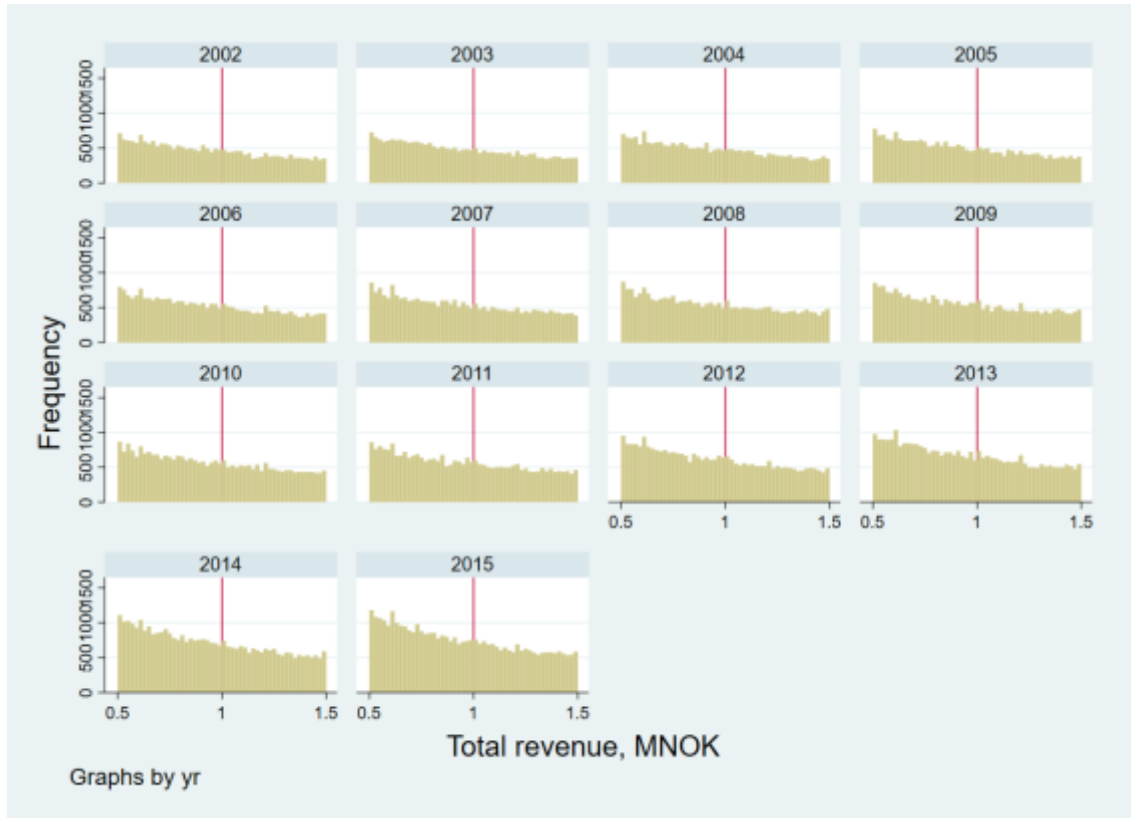
C.4 Annual VAT-reporting 2002-2015 bin width 1000 NOK
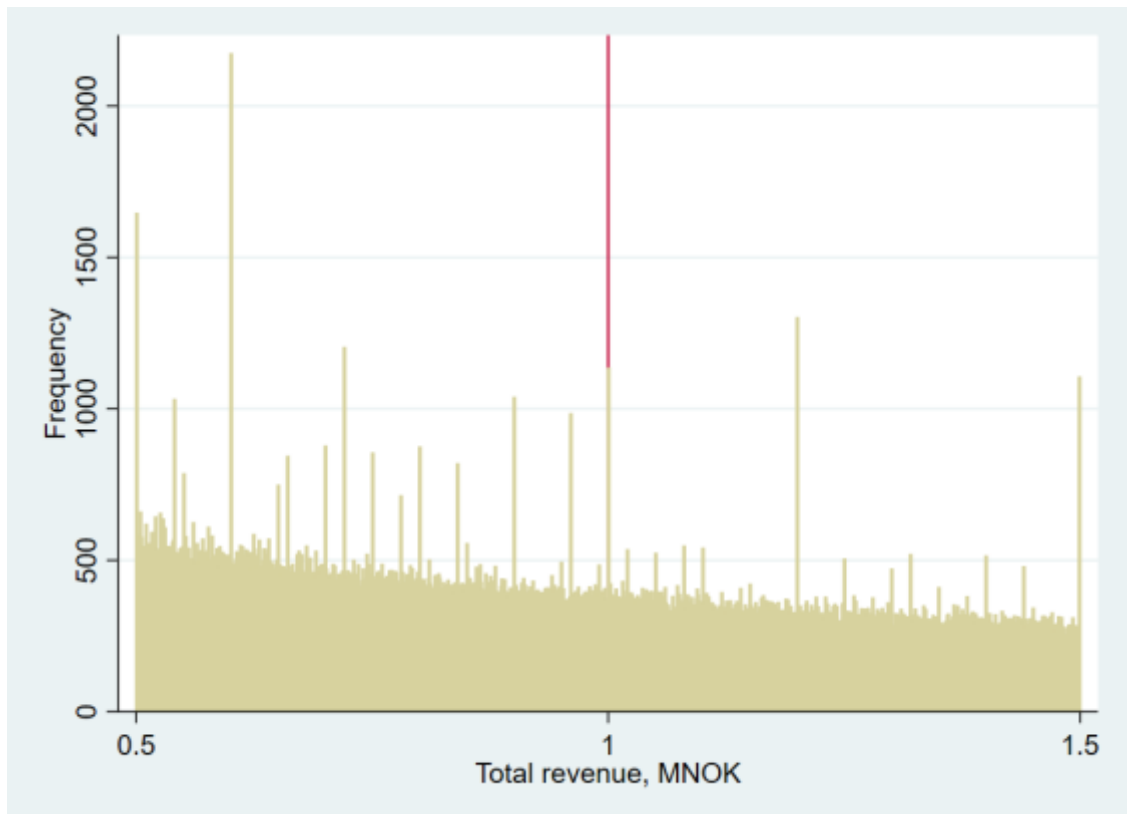
Appendix D: Counterfactual estimation

```
. reg h x x2 x3 x4 x5 x6 x7 if x<4850000 | x>5250000
note: x omitted because of collinearity
```

Source	SS	df	MS			
				Number of obs	=	72
				F(6, 65)	=	1291.72
Model	7757207.27	6	1292867.88	Prob > F	=	0.0000
Residual	65057.8388	65	1000.88983	R-squared	=	0.9917
				Adj R-squared	=	0.9909
Total	7822265.11	71	110172.748	Root MSE	=	31.637

h	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
x	0	(omitted)				
x2	3.38e-09	5.74e-09	0.59	0.558	-8.08e-09	1.48e-08
x3	-2.75e-15	4.00e-15	-0.69	0.494	-1.07e-14	5.24e-15
x4	9.41e-22	1.24e-21	0.76	0.449	-1.53e-21	3.41e-21
x5	-1.65e-28	2.01e-28	-0.82	0.414	-5.66e-28	2.36e-28
x6	1.47e-35	1.67e-35	0.88	0.383	-1.87e-35	4.81e-35
x7	-5.24e-43	5.66e-43	-0.93	0.358	-1.65e-42	6.05e-43
_cons	-214.3675	5966.038	-0.04	0.971	-12129.37	11700.63

Stata regression output from estimation of counterfactual in Figure 19. "h" is number of companies in the representative bin, while x is the level of revenue exponentiated in seven degrees.

Appendix E: Standard Difference

E.1 Detailed SD-statistics:

			Threshold	Bin width	No. of companies in bin -3 to +3	SD +/-1	SD +/-2	SD +/-3
Audit threshold 1994-2010	Total revenue	NOK	5 000 000	50 000	**21 670**	**-0,80**	**0,06**	**-0,22**
					21 332	0,98	1,02	1,36
	Total assets	NOK	20 000 000	200 000	**3 912**	**-0,97**	**0,21**	**0,66**
					3 891	1,89*	2,15*	2,12*
Audit threshold 2012-2015	Total revenue	NOK	5 000 000	50 000	**5 792**	**6,87***	**6,87***	**7,38***
					5 619	-6,78***	-4,19***	-3,56***
	Total assets	NOK	20 000 000	200 000	**1 646**	**1,67***	**0,62**	**0,75**
					1 591	-1,78*	-1,02	-0,82
Audit threshold 2012-2015 with auditor	Total revenue	NOK	5 000 000	50 000	**3 717**	**0,28**	**0,82**	**1,60**
					3 747	-0,52	-0,61	-0,79
	Total assets	NOK	20 000 000	200 000	**1 415**	**1,84***	**0,61**	**0,99**
					1 384	-1,59	-0,48	-0,39
Audit threshold 2012-2015 without auditor	Total revenue	NOK	5 000 000	50 000	**2 075**	**9,60***	**9,04***	**8,99***
					1 872	-13,32***	-7,71***	-6,05***
	Total assets	NOK	20 000 000	200 000	**231**	**-0,14**	**0,14**	**-0,51**
					207	-0,85	-1,70***	-1,36
Audit threshold 2012-2015 without auditor with external accountant	Total revenue	NOK	5 000 000	50 000	**2 435**	**7,65***	**7,88***	**8,00***
					2 279	-8,80	-5,03	-3,50
	Total assets	NOK	20 000 000	200 000	**61**	**N/A**	**N/A**	**N/A**
					52	N/A	N/A	N/A
Audit threshold 2012-2015 without auditor without external accountant	Total revenue	NOK	5 000 000	50 000	**516**	**3,65***	**3,69***	**4,73***
					489	-2,95**	-0,59	-0,91
	Total assets	NOK	20 000 000	200 000	**53**	**N/A**	**N/A**	**N/A**
					47	N/A	N/A	N/A
Simplified VAT reporting	Total revenue	NOK	1 000 000	20 000	**55 031**	**-5,35**	**0,52**	**-0,10**
					54 299	6,96	4,12	7,80
Small companies (2006-2009)	Total revenue	NOK	60 000 000	400 000	**1 126**	**-0,53**	**-0,16**	**-0,89**
					1 091	0,83	0,44	0,76
	Total assets	NOK	30 000 000	200 000	**2 475**	**-1,53**	**-0,86**	**0,13**
					2 479	2,13	2,52	1,64
Small companies (2011-2015)	Total revenue	NOK	70 000 000	400 000	**1 215**	**-0,71**	**-1,04**	**-1,55**
					1 188	1,21	0,91	0,48
	Total assets	NOK	35 000 000	200 000	**3 140**	**0,93**	**1,12**	**1,79***
					3 123	-0,46	1,04	0,21

Standardized difference-tests for the bin just below the threshold. SD +/-1, SD +/-2 and SD +/-3 refer to the standardized difference of a bin relative to the closest, second closest and third closest pair of neighboring bins. The statistically significance on 0.05, 0.01 and 0.001 levels are denoted with *, ** and ***, respectively. Values for the bin just below the threshold are presented in bold, the values for the bin just above the threshold are presented in grey.

E.2 Standardized difference test statistic for the +/-1 bins around the revenue threshold for audit. Companies without auditor (left) and with auditor (right) in the period of 2012 - 2015. Bin width 50,000.
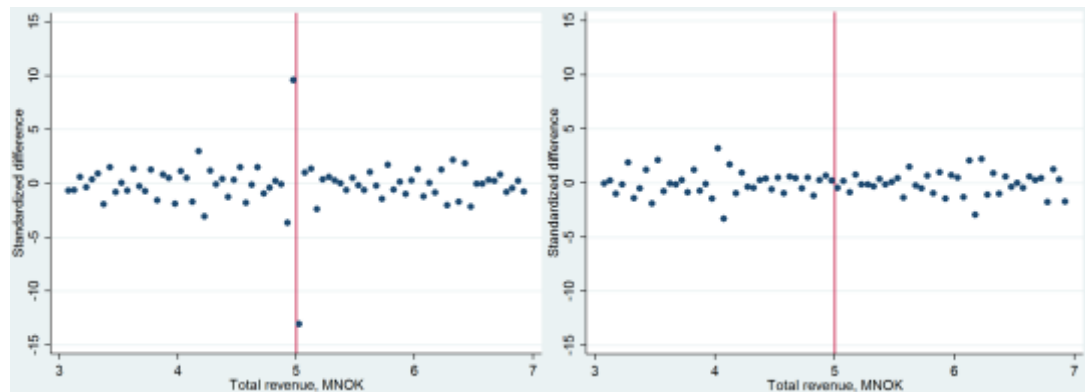
Appendix F: Average bunching response

Bin relativ h	Actual count	Counterfactual count counterfactual	Difference	Proportion	(Calc)	Ave, number of bins managed	Weighted-average bins managed	
-3	885	840,7283	44,2717	0,11	0,34			
-2	919	827,0978	91,9022	0,23	0,47			
-1	1 070	813,5871	256,4129	0,65	0,65			
1	674	800,2238	-126,2238			1,46	0,47	
2	706	786,9207	-80,9207			2,46	0,51	
3	693	773,8123	-80,8123			3,46	0,71	
4	685	760,7562	-75,7562			4,46	0,86	
5	681	747,8607 -	28,8738			5,46	0,40	
6			0			6,46	0,00	
7			0			7,46	0,00	
8			0			8,46	0,00	
9			0			9,46	0,00	
10			0			10,46	0,00	
11			0			11,46	0,00	
12						12,46	0,00	
	6 313	6 351	7,1054E-15	1,00	1,46	**Bins managed**	2,95	
						NOK managed	147 527	57 917 075
						Bin size	50 000	

Average bunching response by companies at the revenue threshold for audit. Method as in Bernard, Burgstahler & Kaya (2017). The difference between the actual and counterfactual in bin 5 is set to 28,87 to make the missing mass equal to the excess mass).

Appendix G: Differences between audited and unaudited companies

	Mean		Median	
	Audited	Unaudited	Audited	Unaudited
Age	12,55	7,55	9	4
Total assets	3 229 000	2 470 000	2 280 000	1 840 000
Growth rate	12,8 %	54,1 %	0,7 %	16,2 %

Calculated in the interval 4.5 to 5.5 mNOK in revenue.

Appendix H: STATA Do-File

```
---------------------------------------------------
*Thresholds.do
*2017-13-07
*Knut Østtveit Aardal, Just Chr. Heide, Norwegian Business School, Nydalen
*---------------------------------------------------
* Program Setup
*---------------------------------------------------
version 14
set more off
clear all
set linesize 80
*---------------------------------------------------
import delimited using KJ2.csv, delimiter(";")


********** ADD VARIABLES **********
* Law firms
* Companies under supervision of the Financial Supervisory Authority
rename    ïcid    cid
merge    1:1 cid yr using "(aardal stokke) konsreg advokatreg mor datter.dta"
drop    if _merge==2
sort    cid yr


xtset    cid yr, yearly

replace i_advokatreg=L.i_advokatreg if yr==2015
replace i_advokatreg=F.i_advokatreg if yr<2000
replace i_advokatreg=F.i_advokatreg if yr<2000
replace i_advokatreg=F.i_advokatreg if yr<2000
replace i_advokatreg=F.i_advokatreg if yr<2000
replace i_advokatreg=F.i_advokatreg if yr<2000
replace i_advokatreg=F.i_advokatreg if yr<2000
```

```
replace i_konsreg=L.i_konsreg if yr==2015
replace i_konsreg=F.i_konsreg if yr<2000
replace i_konsreg=F.i_konsreg if yr<2000
replace i_konsreg=F.i_konsreg if yr<2000
replace i_konsreg=F.i_konsreg if yr<2000
replace i_konsreg=F.i_konsreg if yr<2000
replace i_konsreg=F.i_konsreg if yr<2000


********** GENERATE AND MODIFY VARIABLES **********
rename item_6        ent_type
rename item_11        tot_revenue
rename item_63        nc_assets
rename item_78        c_assets
rename item_87        tot_equity
rename item_91        provisions
rename item_98        oth_liab
rename item_109        tot_c_liab
rename item_13410    aud_cid
rename item_13411    aud_name
rename item_13420    age
rename item_50109    emp_2006_string
rename item_14503    mor
rename item_14504    datter
rename item_50108    nace_noff
rename item_19        op_profit


destring item_13,  replace
destring item_14,  replace
destring item_114, replace
gen wages        = -item_14
gen mgmt_wages    = item_114
gen expenses     = -(item_13 + item_14 + item_18)
gen all_expenses= -(item_13 + item_14 + item_15 + item_16 + item_17 + item_18)
```

```
drop item*
drop if ent_type != "AS"


label var cid           "Company ID"
label var ent_type       "Enterprise type"
label var tot_revenue     "Total revenue"
label var nc_assets       "Non-current assets"
label var c_assets        "Current assets"
label var aud_cid         "Auditor's company ID, string"
label var aud_name        "Auditor's name"


destring nc_assets, replace
destring c_assets, replace
destring tot_equity, replace
destring tot_c_liab, replace
destring aud_cid, replace
destring age, replace
destring emp_2006_string, generate(employees) force


gen double tot_nc_liab    = provisions + oth_liab
gen double tot_liab    = tot_c_liab + tot_nc_liab
gen double tot_assets    = c_assets + nc_assets
gen double eq_liab       = tot_equity + tot_liab
gen has_aud          = 1
replace has_aud       = 0 if missing(aud_cid) | aud_cid==0


label var tot_assets    "Total assets"
label var aud_cid       "Auditor's company ID"
label var has_aud       "Dummy, 1 if company has auditor"


*Separate double NACE-codes
gen nace_1       = ""
```

```
gen nace_2        = ""
replace nace_1    = substr(nace_noff, 1, 6)
replace nace_2    = substr(nace_noff, 7, 6)
replace nace_2    = "" if nace_1==nace_2


gen nace_11=substr(nace_1,1,2)
gen nace_22=substr(nace_2,1,2)
destring nace_11, replace
destring nace_22, replace


********** EXCLUDING COMPANIES THAT ARE MANDATED AUDIT *****
*exclude companies under supervision of the Financial Supervisory Authority
drop if i_konsreg == 1
*exclude law firms
drop if i_advokatreg == 1
*exclude entrepreneurs under the Lottery Act
drop if nace_1=="92.000" & yr>2007
drop if nace_2=="92.710" & yr<2008
*exclude pharmacies and wholesalers
drop if (nace_1=="47.730" | nace_1=="46.460") & yr>2007
drop if (nace_2=="47.730" | nace_2=="46.460") & yr>2007
drop if (nace_1=="52.310" | nace_1=="51.460") & yr<2008
drop if (nace_2=="52.310" | nace_2=="51.460") & yr<2008
*exclude parent companies
drop if mor=="1"
*exclude subsidiaries
drop if datter=="1"


********** DESCRIPTIVE STATISTICS **********
tab yr if yr>2011 & tot_rev<5000000
tab yr if yr>2011 & tot_rev<5000000 & tot_assets>20000000
tab yr if yr>2011 & tot_rev<5000000    & !missing(employees)
tab yr if yr>2011 & tot_rev<5000000    & !missing(employees) & employees>10
```

********** HISTOGRAMS AND SD-TEST OF REVENUE **********

*Revenue distribution all companies

hist tot_revenue if tot_revenue<20000000 & tot_revenue>500000, frequency

gen ln_rev= ln(tot_rev)

hist ln_rev

*Revenue distribution 1994-2010, all companies

```
hist    tot_revenue if (tot_revenue<=7000000 & tot_revenue>=3000000)      ///
     & tot_assets<20000000 & yr<2011,                      ///
     width(50000)                               ///
     xline(5000000)                                ///
     xlabel(3000000 "3" 4000000 "4" 5000000 "5" 6000000 "6" 7000000 "7")    ///
     xtitle("Total revenue, MNOK")                       ///
     frequency
```

*SD-test

```
  drop x h num1 num2 num3 denom SD1 SD2 SD3
  twoway__histogram_gen tot_revenue if (tot_revenue<=7000000          ///
     & tot_revenue>=3000000)                        ///
     & tot_assets<20000000 & yr<2011,                   ///
     width(50000)                             ///
     frequency gen(h x)
  gen num1 = (h - 0.5*(h[_n - 1] + h[_n + 1]))
  gen num2 = (h - 0.5*(h[_n - 2] + h[_n + 2]))
  gen num3 = (h - 0.5*(h[_n - 3] + h[_n + 3]))
  gen denom = sqrt((3/2)*h)
  gen SD1 = num1/denom
  gen SD2 = num2/denom
  gen SD3 = num3/denom
```

*Revenue distribution 2012-2015, all companies

hist   tot_revenue if (tot_revenue<=7000000 & tot_revenue>=3000000)   ///

```
        & tot_assets<20000000 & yr>2011,                    ///
        width(50000)                              ///
        xline(5000000)                               ///
        xlabel(3000000 "3" 4000000 "4" 5000000 "5" 6000000 "6" 7000000 "7")    ///
        xtitle("Total revenue, MNOK")                     ///
        frequency


*SD-test
   twoway__histogram_gen tot_revenue if (tot_revenue<=7000000          ///
        & tot_revenue>=3000000)                       ///
        & tot_assets<20000000 & yr>2011,                 ///
        width(50000)                            ///
        frequency gen(h x)


   gen num1 = (h - 0.5*(h[_n - 1] + h[_n + 1]))
   gen num2 = (h - 0.5*(h[_n - 2] + h[_n + 2]))
   gen num3 = (h - 0.5*(h[_n - 3] + h[_n + 3]))
   gen denom = sqrt((3/2)*h)
   gen SD1 = num1/denom
   gen SD2 = num2/denom
   gen SD3 = num3/denom


*Revenue distribution 2012-2015, companies with auditor
hist    tot_revenue if (tot_revenue<=7000000 & tot_revenue>=3000000)      ///
        & tot_assets<20000000 & yr>2011 & has_aud==1,              ///
        width(50000)                            ///
        xline(5000000)                             ///
        xlabel(3000000 "3" 4000000 "4" 5000000 "5" 6000000 "6" 7000000 "7")    ///
        xtitle("Total revenue, MNOK")                   ///
        frequency


*SD-test
   twoway__histogram_gen tot_revenue if (tot_revenue<=7000000          ///
```

```
            & tot_revenue>=3000000)                          ///
            & tot_assets<20000000 & yr>2011 & has_aud==1,          ///
            width(50000)                              ///
            frequency gen(h x)
            gen num1 = (h - 0.5*(h[_n - 1] + h[_n + 1]))
            gen num2 = (h - 0.5*(h[_n - 2] + h[_n + 2]))
            gen num3 = (h - 0.5*(h[_n - 3] + h[_n + 3]))
            gen denom = sqrt((3/2)*h)
            gen SD1 = num1/denom
            gen SD2 = num2/denom
            gen SD3 = num3/denom


*Revenue distribution 2012-2015, companies without auditor
hist   tot_revenue if (tot_revenue<=7000000 & tot_revenue>=3000000)       ///
            & tot_assets<20000000 & yr>2011 & has_aud==0,          ///
            width(50000)                         ///
            xline(5000000)                              ///
            xlabel(3000000 "3" 4000000 "4" 5000000 "5" 6000000 "6" 7000000 "7")   ///
            xtitle("Total revenue, MNOK")                   ///
            frequency


*SD-test
  twoway__histogram_gen tot_revenue if (tot_revenue<=7000000           ///
            & tot_revenue>=3000000)                     ///
            & tot_assets<20000000 & yr>2011 & has_aud==0,          ///
            width(50000)                          ///
            frequency gen(h x)
            gen num1 = (h - 0.5*(h[_n - 1] + h[_n + 1]))
            gen num2 = (h - 0.5*(h[_n - 2] + h[_n + 2]))
            gen num3 = (h - 0.5*(h[_n - 3] + h[_n + 3]))
            gen denom = sqrt((3/2)*h)
            gen SD1 = num1/denom
            gen SD2 = num2/denom
```

```
gen SD3 = num3/denom
```

*Revenue distribution 2012-2015, all companies by year

```
hist   tot_revenue if (tot_revenue<=7000000 & tot_revenue>=3000000)      ///
       & tot_assets<20000000 & yr>2009, by(yr)                           ///
       width(50000)                                                      ///
       xline(5000000)                                                    ///
       xlabel(3000000 "3" 4000000 "4" 5000000 "5" 6000000 "6" 7000000 "7")   ///
       xtitle("Total revenue, MNOK")                                     ///
       frequency
```

*Revenue distribution 2012-2015, companies with auditor by year

```
hist   tot_revenue if (tot_revenue<=7000000 & tot_revenue>=3000000)      ///
       & tot_assets<20000000 & yr>2011 & has_aud==1,                     ///
       width(50000)                                                      ///
       xline(5000000)                                                    ///
       xlabel(3000000 "3" 4000000 "4" 5000000 "5" 6000000 "6" 7000000 "7")   ///
       xtitle("Total revenue, MNOK")                                     ///
       frequency
```

*SD-test

```
  twoway__histogram_gen tot_revenue if (tot_revenue<=7000000            ///
       & tot_revenue>=3000000)                                           ///
       & tot_assets<20000000 & yr>2011 & has_aud==1,                     ///
       width(50000)                                                      ///
       frequency gen(h x)
       gen num1 = (h - 0.5*(h[_n - 1] + h[_n + 1]))
       gen num2 = (h - 0.5*(h[_n - 2] + h[_n + 2]))
       gen num3 = (h - 0.5*(h[_n - 3] + h[_n + 3]))
       gen denom = sqrt((3/2)*h)
       gen SD1 = num1/denom
       gen SD2 = num2/denom
       gen SD3 = num3/denom'
```

********** CHARACTERISTICS OF AUDITED AND   ///

// UNAUDITED COMPANIES **********

sum age if tot_rev>4500000 & tot_rev<5500000 & tot_assets<20000000   ///

   & yr>2011 & has_aud==1, detail

sum age if tot_rev>4500000 & tot_rev<5500000 & tot_assets<20000000   ///

   & yr>2011 & has_aud==0, detail

sum tot_assets if tot_rev>4500000 & tot_rev<5500000 & tot_assets<20000000   ///

   & yr>2011 & has_aud==1, detail

sum tot_assets if tot_rev>4500000 & tot_rev<5500000 & tot_assets<20000000   ///

   & yr>2011 & has_aud==0, detail

sum growthrate if tot_rev>4500000 & tot_rev<5500000 & tot_assets<20000000   ///

   & yr>2011 & has_aud==1 & growthrate>=-1 & growthrate<10, detail

sum growthrate if tot_rev>4500000 & tot_rev<5500000 & tot_assets<20000000   ///

   & yr>2011 & has_aud==0 & growthrate>=-1 & growthrate<10, detail

********** HISTOGRAMS AND SD-TEST OF TOTAL ASSETS **********

*Total assets distribution 1994-2010, all companies

hist   tot_assets if (tot_assets<=25000000 & tot_assets>=15000000)   ///

   & tot_revenue<5000000 & yr<2011,   ///

   width(200000)   ///

   xline(20000000)   ///

   xlabel(15000000 "15" 20000000 "20" 25000000 "25")   ///

   xtitle("Total assets, MNOK")   ///

   frequency

*SD-test

   drop x h num1 num2 num3 denom SD1 SD2 SD3

   twoway__histogram_gen tot_assets if (tot_assets<=25000000   ///

   & tot_assets>=15000000)   ///

   & tot_revenue<5000000 & yr<2011,   ///

   width(200000)   ///

   frequency gen(h x)

gen num1 = (h - 0.5*(h[_n - 1] + h[_n + 1]))

gen num2 = (h - 0.5*(h[_n - 2] + h[_n + 2]))

gen num3 = (h - 0.5*(h[_n - 3] + h[_n + 3]))

gen denom = sqrt((3/2)*h)

gen SD1 = num1/denom

gen SD2 = num2/denom

gen SD3 = num3/denom

*Total assets distribution 2012-2015, all companies

hist   tot_assets if (tot_assets<=25000000 & tot_assets>=15000000)   ///

    & tot_revenue<5000000 & yr>2011,   ///

    width(200000)   ///

    xline(20000000)   ///

    xlabel(15000000 "15" 20000000 "20" 25000000 "25")   ///

    xtitle("Total assets, MNOK")   ///

    frequency

*SD-test

  drop x h num1 num2 num3 denom SD1 SD2 SD3

  twoway__histogram_gen tot_assets if (tot_assets<=25000000   ///

    & tot_assets>=15000000)   ///

    & tot_revenue<5000000 & yr>2011,   ///

    width(200000)   ///

    frequency gen(h x)

    gen num1 = (h - 0.5*(h[_n - 1] + h[_n + 1]))

    gen num2 = (h - 0.5*(h[_n - 2] + h[_n + 2]))

    gen num3 = (h - 0.5*(h[_n - 3] + h[_n + 3]))

    gen denom = sqrt((3/2)*h)

    gen SD1 = num1/denom

    gen SD2 = num2/denom

    gen SD3 = num3/denom

*Total assets distribution 2012-2015, companies without auditor

```
hist   tot_assets if (tot_assets<=25000000 & tot_assets>=15000000)        ///
       & tot_revenue<5000000 & yr>2011 & has_aud==0,               ///
       width(200000)                              ///
       xline(20000000)                               ///
       xlabel(15000000 "15" 20000000 "20" 25000000 "25")          ///
       xtitle("Total assets, MNOK")                      ///
       frequency


*SD-test
  drop x h num1 num2 num3 denom SD1 SD2 SD3
  twoway__histogram_gen tot_assets if (tot_assets<=25000000         ///
       & tot_assets>=15000000)                          ///
       & tot_revenue<5000000 & yr>2011 & has_aud==0,               ///
       width(200000)                              ///
       frequency gen(h x)
       gen num1 = (h - 0.5*(h[_n - 1] + h[_n + 1]))
       gen num2 = (h - 0.5*(h[_n - 2] + h[_n + 2]))
       gen num3 = (h - 0.5*(h[_n - 3] + h[_n + 3]))
       gen denom = sqrt((3/2)*h)
       gen SD1 = num1/denom
       gen SD2 = num2/denom
       gen SD3 = num3/denom


*Total assets distribution 2012-2015, companies with auditor
hist   tot_assets if (tot_assets<=25000000 & tot_assets>=15000000)        ///
       & tot_revenue<5000000 & yr>2011 & has_aud==1,               ///
       width(200000)                              ///
       xline(20000000)                               ///
       xlabel(15000000 "15" 20000000 "20" 25000000 "25")          ///
       xtitle("Total assets, MNOK")                      ///
       frequency


*SD-test
```

```
drop x h num1 num2 num3 denom SD1 SD2 SD3
twoway__histogram_gen tot_assets if (tot_assets<=25000000          ///
   & tot_assets>=15000000)                                   ///
   & tot_revenue<5000000 & yr>2011 & has_aud==1,             ///
   width(200000)                                             ///
   frequency gen(h x)
   gen num1 = (h - 0.5*(h[_n - 1] + h[_n + 1]))
   gen num2 = (h - 0.5*(h[_n - 2] + h[_n + 2]))
   gen num3 = (h - 0.5*(h[_n - 3] + h[_n + 3]))
   gen denom = sqrt((3/2)*h)
   gen SD1 = num1/denom
   gen SD2 = num2/denom
   gen SD3 = num3/denom


********** HISTOGRAMS AND SD-TEST OF EMPLOYEES **********
*Employees distribution 1994-2010, all companies
hist   employees if (employees<=17 & employees>=5)              ///
   & tot_revenue<5000000 & tot_assets<30000000 & yr<2011,     ///
   width(1)                                                   ///
   xline(11)                                                  ///
   xlabel(5(1)17)                                             ///
   ylabel(10000 "10 000" 20000 "20 000" 30000 "30 000" 40000 "40 000")   ///
   xtitle(Number of employees)                                ///
   frequency


*SD-test
   drop x h num1 num2 num3 denom SD1 SD2 SD3
   twoway__histogram_gen employees if (employees<=17            ///
      & employees>=5)                                          ///
      & tot_revenue<5000000 & tot_assets<30000000 & yr<2011,   ///
      width(1)                                                 ///
      frequency gen(h x)
      gen num1 = (h - 0.5*(h[_n - 1] + h[_n + 1]))
```

```
gen num2 = (h - 0.5*(h[_n - 2] + h[_n + 2]))
gen num3 = (h - 0.5*(h[_n - 3] + h[_n + 3]))
gen denom = sqrt((3/2)*h)
gen SD1 = num1/denom
gen SD2 = num2/denom
gen SD3 = num3/denom


*Employees distribution 2012-2015, all companies
hist    employees if (employees<=17 & employees>=5)                        ///
        & tot_revenue<5000000 & tot_assets<30000000 & yr>2011,             ///
        width(1)                                        ///
        xline(11)                                       ///
        xlabel(5(1)17)                                   ///
        ylabel(2000 "2 000" 4000 "4 000" 6000 "6 000"  ///
        8000 "8 000" 10000 "10 000") ///
        xtitle(Number of employees)                          ///
        frequency


*SD-test
    drop x h num1 num2 num3 denom SD1 SD2 SD3
    twoway__histogram_gen employees if (employees<=17               ///
        & employees>=5)                                      ///
        & tot_revenue<5000000 & tot_assets<30000000 & yr>2011,          ///
        width(1)                                        ///
        frequency gen(h x)
    gen num1 = (h - 0.5*(h[_n - 1] + h[_n + 1]))
    gen num2 = (h - 0.5*(h[_n - 2] + h[_n + 2]))
    gen num3 = (h - 0.5*(h[_n - 3] + h[_n + 3]))
    gen denom = sqrt((3/2)*h)
    gen SD1 = num1/denom
    gen SD2 = num2/denom
    gen SD3 = num3/denom
```

*Employees distribution 2012-2015, companies without auditor

```
hist   employees if (employees<=17 & employees>=5)                        ///
       & tot_revenue<5000000 & tot_assets<30000000 & yr>2011 & has_aud==0,  ///
       width(1)                                            ///
       xline(11)                                           ///
       xlabel(5(1)17)                                      ///
       xtitle(Number of employees)                         ///
       frequency
```

*SD-test

```
  drop x h num1 num2 num3 denom SD1 SD2 SD3
  twoway__histogram_gen employees if (employees<=17           ///
      & employees>=5)                                      ///
      & tot_revenue<5000000 & tot_assets<30000000 & yr>2011 & has_aud==0,  ///
      width(1)                                             ///
      frequency gen(h x)
      gen num1 = (h - 0.5*(h[_n - 1] + h[_n + 1]))
      gen num2 = (h - 0.5*(h[_n - 2] + h[_n + 2]))
      gen num3 = (h - 0.5*(h[_n - 3] + h[_n + 3]))
      gen denom = sqrt((3/2)*h)
      gen SD1 = num1/denom
      gen SD2 = num2/denom
      gen SD3 = num3/denom
```

*Employees distribution 2012-2015, companies with auditor

```
hist   employees if (employees<=17 & employees>=5)                        ///
       & tot_revenue<5000000 & tot_assets<30000000 & yr>2011 & has_aud==1,  ///
       width(1)                                            ///
       xline(11)                                           ///
       xlabel(5(1)17)                                      ///
       xtitle(Number of employees)                         ///
       frequency
```

*SD-test

  drop x h num1 num2 num3 denom SD1 SD2 SD3

  twoway__histogram_gen employees if (employees<=17   ///

    & employees>=5)   ///

    & tot_revenue<5000000 & tot_assets<30000000 & yr>2011 & has_aud==1, ///

    width(1)   ///

    frequency gen(h x)

    gen num1 = (h - 0.5*(h[_n - 1] + h[_n + 1]))

    gen num2 = (h - 0.5*(h[_n - 2] + h[_n + 2]))

    gen num3 = (h - 0.5*(h[_n - 3] + h[_n + 3]))

    gen denom = sqrt((3/2)*h)

    gen SD1 = num1/denom

    gen SD2 = num2/denom

    gen SD3 = num3/denom

*--

*UPLOAD THE DATASET AGAIN TO INCLUDE THE EXCLUDED

// COMPANIES

*Run the commands for the following sections:

// "ADD VARIABLES" and "GENERATE AND MODIFY VARIABLES"

*--

********** HISTOGRAMS OTHER THRESHOLDS **********

********** VAT-REPORTING **********

*Revenue distribution 2002-2015, all companies

hist   tot_revenue if (tot_revenue<=1500000 & tot_revenue>=500000)   ///

    & yr>2001,   ///

    width(20000)   ///

    xline(1000000)   ///

    xlabel(500000 "0.5" 1000000 "1" 1500000 "1.5")   ///

    xtitle("Total revenue, MNOK")   ///

    frequency

```
*SD-test
  drop x h num1 num2 num3 denom SD1 SD2 SD3
  twoway__histogram_gen tot_revenue if (tot_revenue<=1500000          ///
    & tot_revenue>=500000)                                     ///
    & yr>2001,                                            ///
    width(20000)                                        ///
    frequency gen(h x)
    gen num1 = (h - 0.5*(h[_n - 1] + h[_n + 1]))
    gen num2 = (h - 0.5*(h[_n - 2] + h[_n + 2]))
    gen num3 = (h - 0.5*(h[_n - 3] + h[_n + 3]))
    gen denom = sqrt((3/2)*h)
    gen SD1 = num1/denom
    gen SD2 = num2/denom
    gen SD3 = num3/denom
  scatter SD1 x, ylabel(-12 (4) 12) yline(0) xline(1000000)


*Revenue distribution 2002-2015, all companies by year
hist   tot_revenue if (tot_revenue<=1500000 & tot_revenue>=500000)       ///
    & yr>2001,    by(yr)                              ///
    width(20000)                                     ///
    xline(1000000)                                    ///
    xlabel(500000 "0.5" 1000000 "1" 1500000 "1.5")               ///
    xtitle("Total revenue, MNOK")                          ///
    frequency


********** SMALL COMPANY CLASSIFICATION **********
*Revenue distribution 2011-2015, all companies
hist   tot_revenue if (tot_revenue<=80000000 & tot_revenue>=60000000)       ///
    & (tot_assets<35000000 | employees<50) & yr>2010,            ///
    width(400000)                                    ///
    xline(70000000)                                   ///
    xlabel(60000000 "60" 70000000 "70" 80000000 "80")             ///
    xtitle("Total revenue, MNOK")                          ///
```

```
        frequency


*SD-test
  twoway__histogram_gen tot_revenue if (tot_revenue<=80000000          ///
    & tot_revenue>=60000000)                                 ///
    & (tot_assets<35000000 | employees<50) & yr>2010,           ///
    width(400000)                                   ///
    frequency gen(h x)
    gen num1 = (h - 0.5*(h[_n - 1] + h[_n + 1]))
    gen num2 = (h - 0.5*(h[_n - 2] + h[_n + 2]))
    gen num3 = (h - 0.5*(h[_n - 3] + h[_n + 3]))
    gen denom = sqrt((3/2)*h)
    gen SD1 = num1/denom
    gen SD2 = num2/denom
    gen SD3 = num3/denom


*Total assets distribution 2011-2015, all companies
hist   tot_assets if (tot_assets<=40000000 & tot_assets>=30000000)       ///
    & (tot_revenue<70000000 | employees<50) & yr>2010,           ///
    width(200000)                                ///
    xline(35000000)                              ///
    xlabel(30000000 "30" 35000000 "35" 40000000 "40")            ///
    xtitle("Total assets, MNOK")                      ///
    frequency


*SD-test
  drop x h num1 num2 num3 denom SD1 SD2 SD3
  twoway__histogram_gen tot_assets if (tot_assets<=40000000          ///
    & (tot_revenue<70000000 | employees<50) & tot_assets>=30000000)     ///
    & yr>2010,                                ///
    width(200000)                              ///
    frequency gen(h x)
    gen num1 = (h - 0.5*(h[_n - 1] + h[_n + 1]))
```

```
gen num2 = (h - 0.5*(h[_n - 2] + h[_n + 2]))
gen num3 = (h - 0.5*(h[_n - 3] + h[_n + 3]))
gen denom = sqrt((3/2)*h)
gen SD1 = num1/denom
gen SD2 = num2/denom
gen SD3 = num3/denom
```

```
*Employees distribution 2011-2015, all companies
hist   employees if (employees<66 & employees>=35)                        ///
       & (tot_revenue<70000000 | tot_assets<35000000) & yr>2010,          ///
       width(1)                                                           ///
       xline(50)                                                          ///
       xlabel(40 "40" 50 "50" 60 "60")                                    ///
       xtitle(Number of employees)                                        ///
       frequency
```

```
*SD-test
  drop x h num1 num2 num3 denom SD1 SD2 SD3
  twoway__histogram_gen employees if (employees<66                        ///
       & (tot_revenue<70000000 | tot_assets<35000000) & employees>=35)    ///
       & yr>2010,                                                         ///
       width(1)                                                           ///
       frequency gen(h x)
  gen num1 = (h - 0.5*(h[_n - 1] + h[_n + 1]))
  gen num2 = (h - 0.5*(h[_n - 2] + h[_n + 2]))
  gen num3 = (h - 0.5*(h[_n - 3] + h[_n + 3]))
  gen denom = sqrt((3/2)*h)
  gen SD1 = num1/denom
  gen SD2 = num2/denom
  gen SD3 = num3/denom
```

```
*Revenue distribution 2006-2009, all companies
hist   tot_revenue if (tot_revenue<=70000000 & tot_revenue>=50000000)     ///
```

67

```
        & (tot_assets<30000000 | employees<50) & yr>2005 & yr<2010,        ///
        width(400000)                                    ///
        xline(60000000)                                    ///
        xlabel(50000000 "50" 60000000 "60" 70000000 "70")            ///
        xtitle("Total revenue, MNOK")                        ///
        frequency


*SD-test
    twoway__histogram_gen tot_revenue if (tot_revenue<=70000000        ///
        & tot_revenue>=50000000)                            ///
        & (tot_assets<30000000 | employees<50) & yr>2005 & yr<2010,        ///
        width(400000)                                    ///
        frequency gen(h x)
    gen num1 = (h - 0.5*(h[_n - 1] + h[_n + 1]))
    gen num2 = (h - 0.5*(h[_n - 2] + h[_n + 2]))
    gen num3 = (h - 0.5*(h[_n - 3] + h[_n + 3]))
    gen denom = sqrt((3/2)*h)
    gen SD1 = num1/denom
    gen SD2 = num2/denom
    gen SD3 = num3/denom


*Total assets distribution 2006-2009, all companies
hist    tot_assets if (tot_assets<=35000000 & tot_assets>=25000000)        ///
        & (tot_revenue<60000000 | employees<50) & yr>2005 & yr<2010,    ///
        width(200000)                                    ///
        xline(30000000)                                    ///
        xlabel(25000000 "25" 30000000 "30" 35000000 "35")            ///
        xtitle("Total assets, MNOK")                        ///
        frequency


*SD-test
    drop x h num1 num2 num3 denom SD1 SD2 SD3
    twoway__histogram_gen tot_assets if (tot_assets<=35000000            ///
```

```
                & (tot_revenue<60000000 | employees<50) & tot_assets>=25000000)      ///
                & yr>2005 & yr<2010,                                    ///
                width(200000)                                           ///
                frequency gen(h x)
                gen num1 = (h - 0.5*(h[_n - 1] + h[_n + 1]))
                gen num2 = (h - 0.5*(h[_n - 2] + h[_n + 2]))
                gen num3 = (h - 0.5*(h[_n - 3] + h[_n + 3]))
                gen denom = sqrt((3/2)*h)
                gen SD1 = num1/denom
                gen SD2 = num2/denom
                gen SD3 = num3/denom


*Employees distribution 2006-2009, all companies
hist    employees if (employees<66 & employees>=35)                     ///
                & (tot_revenue<60000000 | tot_assets<30000000) & yr>2005 & yr<2010,    ///
                width(1)    discrete                                    ///
                xline(50)                                               ///
                xlabel(40 "40" 50 "50" 60 "60")                         ///
                xtitle(Number of employees)                             ///
                frequency


*SD-test
   drop x h num1 num2 num3 denom SD1 SD2 SD3
   twoway__histogram_gen employees if (employees<66                     ///
                & (tot_revenue<60000000 | tot_assets<30000000) & employees>=35)      ///
                & yr>2005 & yr<2010,                                    ///
                width(1)                                                ///
                frequency gen(h x)
                gen num1 = (h - 0.5*(h[_n - 1] + h[_n + 1]))
                gen num2 = (h - 0.5*(h[_n - 2] + h[_n + 2]))
                gen num3 = (h - 0.5*(h[_n - 3] + h[_n + 3]))
                gen denom = sqrt((3/2)*h)
                gen SD1 = num1/denom
```

69

```
          gen SD2 = num2/denom
          gen SD3 = num3/denom
```

********** EXCLUDING COMPANIES THAT ARE MANDATED AUDIT *****

*exclude companies under supervision of the Financial Supervisory Authority

drop if i_konsreg == 1

*exclude law firms

drop if i_advokatreg == 1

*exclude entrepreneurs under the Lottery Act

drop if nace_1=="92.000" & yr>2007

drop if nace_2=="92.710" & yr<2008

*exclude pharmacies and wholesalers

drop if (nace_1=="47.730" | nace_1=="46.460") & yr>2007

drop if (nace_2=="47.730" | nace_2=="46.460") & yr>2007

drop if (nace_1=="52.310" | nace_1=="51.460") & yr<2008

drop if (nace_2=="52.310" | nace_2=="51.460") & yr<2008

*exclude parent companies

drop if mor=="1"

*exclude subsidiaries

drop if datter=="1"

********** ESTIMATION OF COUNTERFACTUAL DISTRIBUTION ********

```
twoway__histogram_gen tot_revenue if (tot_revenue<=7000000          ///
   & tot_revenue>=3000000)   & tot_assets<20000000                  ///
   & yr>2011,                                     ///
   width(50000)                                   ///
   frequency gen(h x)
```

*Generate polynomials for the regression

gen x2=x^2

gen x3=x^3

gen x4=x^4

gen x5=x^5

```
gen double x6=x^6
gen double x7=x^7


*Estimation of the counterfactual distribution
reg h x x2 x3 x4 x5 x6 x7 if x<4850000 | x>5250000
predict counterfactual
label var counterfactual "Counterfactual"


*"Utvalg" display the values used in the regression
gen utvalg=e(sample)


*Export h and counterfactual values for each bin to Excel for manual
*calculation of bunching statistics. If missing mass < excess mass,
*iterate the process by expanding the upper interval


*Visualization
scatter h x, connect(direct) msize(zero) xline(4850000 5250000, ///
    lpattern(dash)) || scatter counterfactual x, connect(direct) ///
    lpattern(shortdash) msize(zero) xline(5000000)


********** LOCAL POLYNOMIAL SMOOTHING TESTS **********
*Growth
g ltot_revenue = log(tot_rev)
g lgrowth = D.ltot_revenue/L.ltot_revenue
winsor lgrowth, gen(w_lgrowth) p(0.01)
binscatter lgrowth tot_rev if tot_rev<7000000 & tot_rev>3000000 & yr>2011,    ///
    nquantiles(80) linetype(qfit) rd(5000000)                    ///
    xline(5000000)                              ///
    xlabel(3000000 "3" 4000000 "4" 5000000 "5" 6000000 "6" 7000000 "7")       ///
    xtitle("Total revenue, MNOK")                       ///
    ytitle("Logaritmic growth")


lpoly w_lgrowth tot_rev if tot_rev<7000000 & tot_rev>3000000 & yr>2011,    ///
```

71

```
   noscatter ci xline(5000000) bwidth(50000)                    ///
   xline(5000000)                                               ///
   xlabel(3000000 "3" 4000000 "4" 5000000 "5" 6000000 "6" 7000000 "7")     ///
   xtitle("Total revenue, MNOK")                                ///
   ytitle("Logaritmic growth")


*Expenses
winsor expenses, gen(w_expenses) p(0.01)
lpoly w_expenses tot_rev if tot_rev<7000000 & tot_rev>3000000 & yr>2011,    ///
   noscatter ci xline(5000000) bwidth(50000)                    ///
   xline(5000000)                                               ///
   xlabel(3000000 "3" 4000000 "4" 5000000 "5" 6000000 "6" 7000000 "7")     ///
   xtitle("Total revenue, MNOK")                                ///
   ytitle("Expenses, MNOK")                                     ///
   ylabel(2000000 "2" 3000000 "3" 4000000 "4" 5000000 "5" 6000000 "6")


*Expense margin
gen expenses_margin = expenses/tot_revenue
winsor expenses_margin, gen(w_expenses_margin) p(0.1)
binscatter w_expenses_margin tot_rev if tot_rev<7000000 & tot_rev>3000000     ///
   & yr>2011, nquantiles(80) linetype(qfit) rd(5000000)         ///
   xline(5000000)                                               ///
   xlabel(3000000 "3" 4000000 "4" 5000000 "5" 6000000 "6" 7000000 "7")     ///
   xtitle("Total revenue, MNOK")                                ///
   ytitle("Expenses (% of Total revenue)")


lpoly w_expenses_margin tot_rev if tot_rev<7000000 & tot_rev>2000000 &
yr>2011, ///
   noscatter ci xline(5000000) bwidth(50000)                    ///
   xline(5000000)                                               ///
   xlabel(3000000 "3" 4000000 "4" 5000000 "5" 6000000 "6" 7000000 "7")     ///
   xtitle("Total revenue, MNOK")                                ///
   ytitle("Expenses (% of Total revenue)")
```

72

```
*Wages
winsor wages, gen(w_wages) p(0.01)
lpoly w_wages tot_rev if tot_rev<7000000 & tot_rev>3000000 & yr>2011,        ///
   noscatter ci xline(5000000) bwidth(50000)                     ///
   xline(5000000)                                      ///
   xlabel(3000000 "3" 4000000 "4" 5000000 "5" 6000000 "6" 7000000 "7")      ///
   xtitle("Total revenue, MNOK")                          ///
   ytitle("Wages, MNOK")                               ///
   ylabel(1000000 "1" 1500000 "1.5" 2000000 "2" 2500000 "2.5")


*Wage margin
gen wages_margin = wages/tot_rev
winsor wages_margin, gen(w_wages_margin) p(0.01)
binscatter w_wages_margin tot_rev if tot_rev<7000000 & tot_rev>3000000        ///
   & yr>2011, nquantiles(80) linetype(qfit) rd(5000000)              ///
   xline(5000000)                                      ///
   xlabel(3000000 "3" 4000000 "4" 5000000 "5" 6000000 "6" 7000000 "7")      ///
   xtitle("Total revenue, MNOK")                          ///
   ytitle("Wages (% of Total revenue)")


lpoly w_wages_margin tot_rev if tot_rev<7500000 & tot_rev>2500000 & yr>2011,///
   noscatter ci xline(5000000) bwidth(50000)                     ///
   xline(5000000)                                      ///
   xlabel(3000000 "3" 4000000 "4" 5000000 "5" 6000000 "6" 7000000 "7")      ///
   xtitle("Total revenue, MNOK")                          ///
   ytitle("Wages (% of Total revenue)")


*Operating profit
winsor op_profit, gen(w_op_profit) p(0.01)
lpoly w_op_profit tot_rev if tot_rev<7000000 & tot_rev>3000000 & yr>2011,    ///
   noscatter ci xline(5000000) bwidth(50000)                     ///
   xline(5000000)                                      ///
```

```
        xlabel(3000000 "3" 4000000 "4" 5000000 "5" 6000000 "6" 7000000 "7")      ///
        xtitle("Total revenue, MNOK")                          ///
        ytitle("Operating profit, NOK")                        ///
        ylabel(300000 "300,000" 400000 "400,000" 500000 "500,000" 600000 "600,000")


*Operating profit margin
gen op_profit_margin = op_profit/tot_rev
winsor op_profit_margin, gen(w_op_profit_margin) p(0.01)
binscatter w_op_profit_margin tot_rev if tot_rev<7000000 & tot_rev>3000000    ///
        & yr>2011, nquantiles(80) linetype(qfit) rd(5000000)          ///
        xline(5000000)                                 ///
        xlabel(3000000 "3" 4000000 "4" 5000000 "5" 6000000 "6" 7000000 "7")      ///
        xtitle("Total revenue, MNOK")                          ///
        ytitle("Operating profit (% of Total revenue)")


lpoly w_op_profit_margin tot_rev if tot_rev<7500000 & tot_rev>2500000 ///
        & yr>2011,  ///
        noscatter ci xline(5000000) bwidth(50000)                ///
        xline(5000000)                               ///
        xlabel(3000000 "3" 4000000 "4" 5000000 "5" 6000000 "6" 7000000 "7")      ///
        xtitle("Total revenue, MNOK")                        ///
        ytitle("Operating profit (% of Total revenue)")


* Total equity
winsor tot_equity, gen(w_tot_equity) p(0.01)
lpoly w_tot_equity tot_rev if tot_rev<7000000 & tot_rev>3000000 & yr>2011,    ///
        noscatter ci xline(5000000) bwidth(50000)                ///
        xline(5000000)                             ///
        xlabel(3000000 "3" 4000000 "4" 5000000 "5" 6000000 "6" 7000000 "7")      ///
        xtitle("Total revenue, MNOK")                      ///
        ytitle("Equity, MNOK")                         ///
        ylabel(1000000 "1" 1500000 "1.5" 2000000 "2" 2500000 "2.5")
```

74

```
* Total equity margin
gen tot_equity_margin = tot_equity/tot_rev
winsor tot_equity_margin, gen(w_tot_equity_margin) p(0.01)
binscatter w_tot_equity_margin tot_rev if tot_rev<7000000 & tot_rev>3000000 ///
   & yr>2011, nquantiles(80) linetype(qfit) rd(5000000)              ///
   xline(5000000)                                               ///
   xlabel(3000000 "3" 4000000 "4" 5000000 "5" 6000000 "6" 7000000 "7")    ///
   xtitle("Total revenue, MNOK")                                 ///
   ytitle("Equity (% of Total revenue)")


lpoly w_tot_equity_margin tot_rev if tot_rev<7500000 & tot_rev>2500000 & ///
   yr>2011, ///
   noscatter ci xline(5000000) bwidth(50000)                      ///
   xline(5000000)                                               ///
   xlabel(3000000 "3" 4000000 "4" 5000000 "5" 6000000 "6" 7000000 "7")    ///
   xtitle("Total revenue, MNOK")                                 ///
   ytitle("Equity (% of Total revenue)")


********** PERSISTENCE RATES **********
gen int bin=-1000
replace bin=-5    if tot_revenue>=4250000 & tot_revenue<4400000
replace bin=-4    if tot_revenue>=4400000 & tot_revenue<4550000
replace bin=-3    if tot_revenue>=4550000 & tot_revenue<4700000
replace bin=-2    if tot_revenue>=4700000 & tot_revenue<4850000
replace bin=-1    if tot_revenue>=4850000 & tot_revenue<5000000
replace bin=0     if tot_revenue>=5000000 & tot_revenue<5150000
replace bin=1     if tot_revenue>=5150000 & tot_revenue<5300000
replace bin=2     if tot_revenue>=5300000 & tot_revenue<5450000
replace bin=3     if tot_revenue>=5450000 & tot_revenue<5600000
replace bin=4     if tot_revenue>=5600000 & tot_revenue<5750000


gen persistence_t1=0
gen persistence_t2=0
```

```
gen persistence_t3=0

gen persistence_t4=0

replace persistence_t1=1 if L.bin==bin

replace persistence_t2=1 if L2.bin==bin

replace persistence_t3=1 if L3.bin==bin

replace persistence_t4=1 if L4.bin==bin


*One year

binscatter persistence_t1 bin if bin!=-1000 & yr>2011,                ///
   discrete linetype(connect) xlabel(-5 "-5" -4 "-4" -3 "-3" -2 "-2"       ///
   -1 "-1" 0 "1" 1 "2" 2 "3" 3 "4" 4 "5") xline(-0.5) ylabel(0 (0.02) 0.1) ///
   ytitle("Persistence rate") xtitle("Distance from threshold")


*Two year

binscatter persistence_t2 bin if bin!=-1000 & yr>2011,                ///
   discrete linetype(connect) xlabel(-5 "-5" -4 "-4" -3 "-3" -2 "-2"       ///
   -1 "-1" 0 "1" 1 "2" 2 "3" 3 "4" 4 "5") xline(-0.5) ylabel(0 (0.02) 0.1) ///
   ytitle("Persistence rate") xtitle("Distance from threshold")


*Three year

binscatter persistence_t3 bin if bin!=-1000 & yr>2011,                ///
   discrete linetype(connect) xlabel(-5 "-5" -4 "-4" -3 "-3" -2 "-2"       ///
   -1 "-1" 0 "1" 1 "2" 2 "3" 3 "4" 4 "5") xline(-0.5) ylabel(0 (0.02) 0.1) ///
   ytitle("Persistence rate") xtitle("Distance from threshold")


*Four year

binscatter persistence_t4 bin if bin!=-1000 & yr>2011,                ///
   discrete linetype(connect) xlabel(-5 "-5" -4 "-4" -3 "-3" -2 "-2"       ///
   -1 "-1" 0 "1" 1 "2" 2 "3" 3 "4" 4 "5") xline(-0.5) ylabel(0 (0.02) 0.1) ///
   ytitle("Persistence rate") xtitle("Distance from threshold")


*Each year

binscatter persistence_t1 bin if bin!=-1000 & yr>2011, by(yr)          ///
```

76

```
     discrete linetype(connect) xlabel(-5 "-5" -4 "-4" -3 "-3" -2 "-2"        ///
      -1 "-1" 0 "1" 1 "2" 2 "3" 3 "4" 4 "5") xline(-0.5)                ///
      ytitle("Persistence rate") xtitle("Distance from threshold")


     ********** BUNCHING PR. NACE-SECTION **********
     gen nace_11=substr(nace_1,1,2)
     gen nace_22=substr(nace_2,1,2)
     destring nace_11, replace
     destring nace_22, replace


     *NACE-SECTION A:
     hist tot_revenue if (tot_revenue<=7000000 & tot_revenue>=3000000) ///
      & tot_assets<20000000 & yr>2011        ///
      & (      ///
      nace_11==1   | ///
      nace_22==1   | ///
      nace_11==2   | ///
      nace_22==2   | ///
      nace_11==3    | ///
      nace_22==3    ///
      ),    ///
      width(100000)         ///
      xline(5000000)         ///
      xlabel(3000000 "3" 4000000 "4" 5000000 "5" 6000000 "6" 7000000 "7") ///
      xtitle("Total revenue, MNOK")        ///
      frequency


     *NACE-SECTION B:
     hist tot_revenue if (tot_revenue<=7000000 & tot_revenue>=3000000) ///
      & tot_assets<20000000 & yr>2011        ///
      & (      ///
      nace_11==5    | ///
```

```
nace_22==5   | ///
nace_11==6   | ///
nace_22==6   | ///
nace_11==7   | ///
nace_22==7   | ///
nace_11==8   | ///
nace_22==8   | ///
nace_11==9    | ///
nace_22==9    ///
),     ///
width(100000)          ///
xline(5000000)          ///
xlabel(3000000 "3" 4000000 "4" 5000000 "5" 6000000 "6" 7000000 "7") ///
xtitle("Total revenue, MNOK")        ///
frequency
```

*NACE-SECTION C:
```
hist tot_revenue if (tot_revenue<=7000000 & tot_revenue>=3000000)  ///
 & tot_assets<20000000 & yr>2011          ///
 & (      ///
 nace_11==10   | ///
 nace_22==10   | ///
 nace_11==11   | ///
 nace_22==11   | ///
 nace_11==12   | ///
 nace_22==12   | ///
 nace_11==13   | ///
 nace_22==13   | ///
 nace_11==14   | ///
 nace_22==14   | ///
 nace_11==15   | ///
 nace_22==15   | ///
 nace_11==16   | ///
```

78

nace_22==16 | ///

nace_11==17 | ///

nace_22==17 | ///

nace_11==18 | ///

nace_22==18 | ///

nace_11==19 | ///

nace_22==19 | ///

nace_11==20 | ///

nace_22==20 | ///

nace_11==21 | ///

nace_22==21 | ///

nace_11==22 | ///

nace_22==22 | ///

nace_11==23 | ///

nace_22==23 | ///

nace_11==24 | ///

nace_22==24 | ///

nace_11==25 | ///

nace_22==25 | ///

nace_11==26 | ///

nace_22==26 | ///

nace_11==27 | ///

nace_22==27 | ///

nace_11==28 | ///

nace_22==28 | ///

nace_11==29 | ///

nace_22==29 | ///

nace_11==30 | ///

nace_22==30 | ///

nace_11==31 | ///

nace_22==31 | ///

nace_11==32 | ///

nace_22==32 | ///

```
nace_11==33    | ///
nace_22==33    ///
),     ///
width(100000)          ///
xline(5000000)         ///
xlabel(3000000 "3" 4000000 "4" 5000000 "5" 6000000 "6" 7000000 "7") ///
xtitle("Total revenue, MNOK")        ///
frequency


*SD-test
twoway__histogram_gen tot_revenue if (tot_revenue<=7000000    ///
 & tot_revenue>=3000000)          ///
 & tot_assets<20000000 & yr>2011         ///
 & (          ///
nace_11==10   | ///
nace_22==10   | ///
nace_11==11   | ///
nace_22==11   | ///
nace_11==12   | ///
nace_22==12   | ///
nace_11==13   | ///
nace_22==13   | ///
nace_11==14   | ///
nace_22==14   | ///
nace_11==15   | ///
nace_22==15   | ///
nace_11==16   | ///
nace_22==16   | ///
nace_11==17   | ///
nace_22==17   | ///
nace_11==18   | ///
nace_22==18   | ///
nace_11==19   | ///
```

```
nace_22==19    | ///
nace_11==20    | ///
nace_22==20    | ///
nace_11==21    | ///
nace_22==21    | ///
nace_11==22    | ///
nace_22==22    | ///
nace_11==23    | ///
nace_22==23    | ///
nace_11==24    | ///
nace_22==24    | ///
nace_11==25    | ///
nace_22==25    | ///
nace_11==26    | ///
nace_22==26    | ///
nace_11==27    | ///
nace_22==27    | ///
nace_11==28    | ///
nace_22==28    | ///
nace_11==29    | ///
nace_22==29    | ///
nace_11==30    | ///
nace_22==30    | ///
nace_11==31    | ///
nace_22==31    | ///
nace_11==32    | ///
nace_22==32    | ///
nace_11==33    | ///
nace_22==33    ///
),      ///
width(100000)        ///
frequency gen(h x)
gen num1 = (h - 0.5*(h[_n - 1] + h[_n + 1]))
```

```
gen num2 = (h - 0.5*(h[_n - 2] + h[_n + 2]))
gen num3 = (h - 0.5*(h[_n - 3] + h[_n + 3]))
gen denom = sqrt((3/2)*h)
gen SD1 = num1/denom
gen SD2 = num2/denom
gen SD3 = num3/denom


*NACE-SECTION D:
hist tot_revenue if (tot_revenue<=7000000 & tot_revenue>=3000000)  ///
  & tot_assets<20000000 & yr>2011          ///
  & (      ///
  nace_11==35   | ///
  nace_22==35    ///
  ),    ///
  width(100000)         ///
  xline(5000000)         ///
  xlabel(3000000 "3" 4000000 "4" 5000000 "5" 6000000 "6" 7000000 "7") ///
  xtitle("Total revenue, MNOK")       ///
  frequency


*NACE-SECTION E:
hist tot_revenue if (tot_revenue<=7000000 & tot_revenue>=3000000)  ///
  & tot_assets<20000000 & yr>2011          ///
  & (      ///
  nace_11==36   | ///
  nace_22==36   | ///
  nace_11==37   | ///
  nace_22==37   | ///
  nace_11==38   | ///
  nace_22==38    ///
  ),    ///
  width(100000)         ///
  xline(5000000)         ///
```

```
 xlabel(3000000 "3" 4000000 "4" 5000000 "5" 6000000 "6" 7000000 "7") ///
 xtitle("Total revenue, MNOK")        ///
 frequency


*NACE-SECTION F:
hist tot_revenue if (tot_revenue<=7000000 & tot_revenue>=3000000)  ///
 & tot_assets<20000000 & yr>2011        ///
 & (      ///
 nace_11==41  | ///
 nace_22==41  | ///
 nace_11==42  | ///
 nace_22==42  | ///
 nace_11==43  | ///
 nace_22==43   ///
 ),    ///
 width(100000)        ///
 xline(5000000)        ///
 xlabel(3000000 "3" 4000000 "4" 5000000 "5" 6000000 "6" 7000000 "7") ///
 xtitle("Total revenue, MNOK")        ///
 frequency


*SD-test
 drop x h num1 num2 num3 denom SD1 SD2 SD3
 twoway__histogram_gen tot_revenue if (tot_revenue<=7000000     ///
 & tot_revenue>=3000000)          ///
 & tot_assets<20000000 & yr>2011        ///
 & (        ///
 nace_11==41  | ///
 nace_22==41  | ///
 nace_11==42  | ///
 nace_22==42  | ///
 nace_11==43  | ///
 nace_22==43   ///
```

```
),         ///
width(100000)          ///
frequency gen(h x)
gen num1 = (h - 0.5*(h[_n - 1] + h[_n + 1]))
gen num2 = (h - 0.5*(h[_n - 2] + h[_n + 2]))
gen num3 = (h - 0.5*(h[_n - 3] + h[_n + 3]))
gen denom = sqrt((3/2)*h)
gen SD1 = num1/denom
gen SD2 = num2/denom
gen SD3 = num3/denom


*NACE-SECTION G:
hist tot_revenue if (tot_revenue<=7000000 & tot_revenue>=3000000)  ///
 & tot_assets<20000000 & yr>2011          ///
 & (       ///
 nace_11==45   | ///
 nace_22==45   | ///
 nace_11==46   | ///
 nace_22==46   | ///
 nace_11==47   | ///
 nace_22==47    ///
 ),    ///
 width(100000)          ///
 xline(5000000)          ///
 xlabel(3000000 "3" 4000000 "4" 5000000 "5" 6000000 "6" 7000000 "7") ///
 xtitle("Total revenue, MNOK")        ///
 frequency


*SD-test
drop x h num1 num2 num3 denom SD1 SD2 SD3
twoway__histogram_gen tot_revenue if (tot_revenue<=7000000     ///
 & tot_revenue>=3000000)          ///
 & tot_assets<20000000 & yr>2011        ///
```

```
    & (              ///
    nace_11==45   | ///
    nace_22==45   | ///
    nace_11==46   | ///
    nace_22==46   | ///
    nace_11==47   | ///
    nace_22==47     ///
    ),        ///
    width(100000)          ///
    frequency gen(h x)
    gen num1 = (h - 0.5*(h[_n - 1] + h[_n + 1]))
    gen num2 = (h - 0.5*(h[_n - 2] + h[_n + 2]))
    gen num3 = (h - 0.5*(h[_n - 3] + h[_n + 3]))
    gen denom = sqrt((3/2)*h)
    gen SD1 = num1/denom
    gen SD2 = num2/denom
    gen SD3 = num3/denom


*NACE-SECTION H:
hist tot_revenue if (tot_revenue<=7000000 & tot_revenue>=3000000)  ///
    & tot_assets<20000000 & yr>2011          ///
    & (     ///
    nace_11==49   | ///
    nace_22==49   | ///
    nace_11==50   | ///
    nace_22==50   | ///
    nace_11==51   | ///
    nace_22==51   | ///
    nace_11==52   | ///
    nace_22==52   | ///
    nace_11==53   | ///
    nace_22==53     ///
    ),    ///
```

```
    width(100000)            ///
    xline(5000000)           ///
    xlabel(3000000 "3" 4000000 "4" 5000000 "5" 6000000 "6" 7000000 "7") ///
    xtitle("Total revenue, MNOK")        ///
    frequency


*SD-test
drop x h num1 num2 num3 denom SD1 SD2 SD3
twoway__histogram_gen tot_revenue if (tot_revenue<=7000000    ///
 & tot_revenue>=3000000)          ///
 & tot_assets<20000000 & yr>2011         ///
 & (          ///
 nace_11==49  | ///
 nace_22==49  | ///
 nace_11==50  | ///
 nace_22==50  | ///
 nace_11==51  | ///
 nace_22==51  | ///
 nace_11==52  | ///
 nace_22==52  | ///
 nace_11==53  | ///
 nace_22==53   ///
 ),        ///
    width(100000)           ///
    frequency gen(h x)
gen num1 = (h - 0.5*(h[_n - 1] + h[_n + 1]))
gen num2 = (h - 0.5*(h[_n - 2] + h[_n + 2]))
gen num3 = (h - 0.5*(h[_n - 3] + h[_n + 3]))
gen denom = sqrt((3/2)*h)
gen SD1 = num1/denom
gen SD2 = num2/denom
gen SD3 = num3/denom
```

```
*NACE-SECTION I:
hist tot_revenue if (tot_revenue<=7000000 & tot_revenue>=3000000)  ///
 & tot_assets<20000000 & yr>2011          ///
 & (       ///
 nace_11==55   | ///
 nace_22==55   | ///
 nace_11==56   | ///
 nace_22==56    ///
 ),    ///
 width(100000)          ///
 xline(5000000)          ///
 xlabel(3000000 "3" 4000000 "4" 5000000 "5" 6000000 "6" 7000000 "7") ///
 xtitle("Total revenue, MNOK")        ///
 frequency


*SD-test
 drop x h num1 num2 num3 denom SD1 SD2 SD3
 twoway__histogram_gen tot_revenue if (tot_revenue<=7000000     ///
 & tot_revenue>=3000000)          ///
 & tot_assets<20000000 & yr>2011          ///
 & (         ///
 nace_11==55   | ///
 nace_22==55   | ///
 nace_11==56   | ///
 nace_22==56    ///
 ),      ///
 width(100000)          ///
 frequency gen(h x)
 gen num1 = (h - 0.5*(h[_n - 1] + h[_n + 1]))
 gen num2 = (h - 0.5*(h[_n - 2] + h[_n + 2]))
 gen num3 = (h - 0.5*(h[_n - 3] + h[_n + 3]))
 gen denom = sqrt((3/2)*h)
 gen SD1 = num1/denom
```

```
gen SD2 = num2/denom
gen SD3 = num3/denom


*NACE-SECTION J:
hist tot_revenue if (tot_revenue<=7000000 & tot_revenue>=3000000)  ///
 & tot_assets<20000000 & yr>2011        ///
 & (      ///
 nace_11==58  | ///
 nace_22==58  | ///
 nace_11==59  | ///
 nace_22==59  | ///
 nace_11==60  | ///
 nace_22==60  | ///
 nace_11==61  | ///
 nace_22==61  | ///
 nace_11==62  | ///
 nace_22==62  | ///
 nace_11==63  | ///
 nace_22==63   ///
 ),   ///
 width(100000)          ///
 xline(5000000)          ///
 xlabel(3000000 "3" 4000000 "4" 5000000 "5" 6000000 "6" 7000000 "7") ///
 xtitle("Total revenue, MNOK")       ///
 frequency


*SD-test
 drop x h num1 num2 num3 denom SD1 SD2 SD3
 twoway__histogram_gen tot_revenue if (tot_revenue<=7000000    ///
 & tot_revenue>=3000000)       ///
 & tot_assets<20000000 & yr>2011       ///
 & (       ///
 nace_11==58  | ///
```

```
nace_22==58   | ///
nace_11==59   | ///
nace_22==59   | ///
nace_11==60   | ///
nace_22==60   | ///
nace_11==61   | ///
nace_22==61   | ///
nace_11==62   | ///
nace_22==62   | ///
nace_11==63   | ///
nace_22==63    ///
),       ///
width(100000)          ///
frequency gen(h x)
gen num1 = (h - 0.5*(h[_n - 1] + h[_n + 1]))
gen num2 = (h - 0.5*(h[_n - 2] + h[_n + 2]))
gen num3 = (h - 0.5*(h[_n - 3] + h[_n + 3]))
gen denom = sqrt((3/2)*h)
gen SD1 = num1/denom
gen SD2 = num2/denom
gen SD3 = num3/denom


*NACE-SECTION K:
hist tot_revenue if (tot_revenue<=7000000 & tot_revenue>=3000000)  ///
 & tot_assets<20000000 & yr>2011          ///
 & (      ///
 nace_11==64   | ///
 nace_22==64   | ///
 nace_11==65   | ///
 nace_22==65   | ///
 nace_11==66   | ///
 nace_22==66    ///
 ),    ///
```

89

```
  width(100000)              ///
  xline(5000000)             ///
  xlabel(3000000 "3" 4000000 "4" 5000000 "5" 6000000 "6" 7000000 "7") ///
  xtitle("Total revenue, MNOK")          ///
  frequency
```

```
*NACE-SECTION L:
hist tot_revenue if (tot_revenue<=7000000 & tot_revenue>=3000000)  ///
  & tot_assets<20000000 & yr>2011          ///
  & (        ///
  nace_11==68   | ///
  nace_22==68    ///
  ),    ///
  width(100000)             ///
  xline(5000000)             ///
  xlabel(3000000 "3" 4000000 "4" 5000000 "5" 6000000 "6" 7000000 "7") ///
  xtitle("Total revenue, MNOK")          ///
  frequency
```

```
*SD-test
drop x h num1 num2 num3 denom SD1 SD2 SD3
twoway__histogram_gen tot_revenue if (tot_revenue<=7000000     ///
  & tot_revenue>=3000000)           ///
  & tot_assets<20000000 & yr>2011          ///
  & (           ///
  nace_11==68   | ///
  nace_22==68    ///
  ),       ///
  width(100000)           ///
  frequency gen(h x)
gen num1 = (h - 0.5*(h[_n - 1] + h[_n + 1]))
gen num2 = (h - 0.5*(h[_n - 2] + h[_n + 2]))
gen num3 = (h - 0.5*(h[_n - 3] + h[_n + 3]))
```

90

```
gen denom = sqrt((3/2)*h)
gen SD1 = num1/denom
gen SD2 = num2/denom
gen SD3 = num3/denom


*NACE-SECTION M:
hist tot_revenue if (tot_revenue<=7000000 & tot_revenue>=3000000)  ///
 & tot_assets<20000000 & yr>2011          ///
 & (      ///
 nace_11==69  | ///
 nace_22==69  | ///
 nace_11==70  | ///
 nace_22==70  | ///
 nace_11==71  | ///
 nace_22==71  | ///
 nace_11==72  | ///
 nace_22==72  | ///
 nace_11==73  | ///
 nace_22==73  | ///
 nace_11==74  | ///
 nace_22==74  | ///
 nace_11==75  | ///
 nace_22==75   ///
 ),    ///
 width(100000)          ///
 xline(5000000)         ///
 xlabel(3000000 "3" 4000000 "4" 5000000 "5" 6000000 "6" 7000000 "7") ///
 xtitle("Total revenue, MNOK")         ///
 frequency


*SD-test
drop x h num1 num2 num3 denom SD1 SD2 SD3
twoway__histogram_gen tot_revenue if (tot_revenue<=7000000     ///
```

```
                  & tot_revenue>=3000000)           ///
                  & tot_assets<20000000 & yr>2011         ///
                  & (         ///
                  nace_11==69   | ///
                  nace_22==69   | ///
                  nace_11==70   | ///
                  nace_22==70   | ///
                  nace_11==71   | ///
                  nace_22==71   | ///
                  nace_11==72   | ///
                  nace_22==72   | ///
                  nace_11==73   | ///
                  nace_22==73   | ///
                  nace_11==74   | ///
                  nace_22==74   | ///
                  nace_11==75   | ///
                  nace_22==75   ///
                  ),       ///
                  width(100000)          ///
                  frequency gen(h x)
                  gen num1 = (h - 0.5*(h[_n - 1] + h[_n + 1]))
                  gen num2 = (h - 0.5*(h[_n - 2] + h[_n + 2]))
                  gen num3 = (h - 0.5*(h[_n - 3] + h[_n + 3]))
                  gen denom = sqrt((3/2)*h)
                  gen SD1 = num1/denom
                  gen SD2 = num2/denom
                  gen SD3 = num3/denom


*NACE-SECTION N:
hist tot_revenue if (tot_revenue<=7000000 & tot_revenue>=3000000) ///
  & tot_assets<20000000 & yr>2011          ///
  & (      ///
  nace_11==77   | ///
```

```
nace_22==77   | ///
nace_11==78   | ///
nace_22==78   | ///
nace_11==79   | ///
nace_22==79   | ///
nace_11==80   | ///
nace_22==80   | ///
nace_11==81   | ///
nace_22==81   | ///
nace_11==82   | ///
nace_22==82    ///
),     ///
width(100000)          ///
xline(5000000)          ///
xlabel(3000000 "3" 4000000 "4" 5000000 "5" 6000000 "6" 7000000 "7") ///
xtitle("Total revenue, MNOK")        ///
 frequency

*SD-test
drop x h num1 num2 num3 denom SD1 SD2 SD3
twoway__histogram_gen tot_revenue if (tot_revenue<=7000000     ///
 & tot_revenue>=3000000)        ///
 & tot_assets<20000000 & yr>2011        ///
 & (        ///
nace_11==77   | ///
nace_22==77   | ///
nace_11==78   | ///
nace_22==78   | ///
nace_11==79   | ///
nace_22==79   | ///
nace_11==80   | ///
nace_22==80   | ///
nace_11==81   | ///
```

```
nace_22==81   | ///
nace_11==82   | ///
nace_22==82   ///
),        ///
width(100000)          ///
frequency gen(h x)
gen num1 = (h - 0.5*(h[_n - 1] + h[_n + 1]))
gen num2 = (h - 0.5*(h[_n - 2] + h[_n + 2]))
gen num3 = (h - 0.5*(h[_n - 3] + h[_n + 3]))
gen denom = sqrt((3/2)*h)
gen SD1 = num1/denom
gen SD2 = num2/denom
gen SD3 = num3/denom


*NACE-SECTION O:
hist tot_revenue if (tot_revenue<=7000000 & tot_revenue>=3000000)  ///
  & tot_assets<20000000 & yr>2011          ///
  & (      ///
  nace_11==84   | ///
  nace_22==84   ///
  ),   ///
  width(100000)          ///
  xline(5000000)          ///
  xlabel(3000000 "3" 4000000 "4" 5000000 "5" 6000000 "6" 7000000 "7") ///
  xtitle("Total revenue, MNOK")          ///
  frequency


*NACE-SECTION P:
hist tot_revenue if (tot_revenue<=7000000 & tot_revenue>=3000000)  ///
  & tot_assets<20000000 & yr>2011          ///
  & (      ///
  nace_11==85   | ///
  nace_22==85   ///
```

```
),        ///
 width(100000)              ///
 xline(5000000)              ///
 xlabel(3000000 "3" 4000000 "4" 5000000 "5" 6000000 "6" 7000000 "7") ///
 xtitle("Total revenue, MNOK")          ///
 frequency
```

```
*NACE-SECTION Q:
hist tot_revenue if (tot_revenue<=7000000 & tot_revenue>=3000000)  ///
 & tot_assets<20000000 & yr>2011           ///
 & (      ///
 nace_11==86   | ///
 nace_22==86   | ///
 nace_11==87   | ///
 nace_22==87   | ///
 nace_11==88   | ///
 nace_22==88    ///
 ),        ///
 width(100000)              ///
 xline(5000000)              ///
 xlabel(3000000 "3" 4000000 "4" 5000000 "5" 6000000 "6" 7000000 "7") ///
 xtitle("Total revenue, MNOK")          ///
 frequency
```

```
*SD-test
 drop x h num1 num2 num3 denom SD1 SD2 SD3
 twoway__histogram_gen tot_revenue if (tot_revenue<=7000000     ///
 & tot_revenue>=3000000)           ///
 & tot_assets<20000000 & yr>2011           ///
 & (          ///
 nace_11==86   | ///
 nace_22==86   | ///
 nace_11==87   | ///
```

```
nace_22==87   | ///
nace_11==88   | ///
nace_22==88   ///
),       ///
width(100000)          ///
frequency gen(h x)
gen num1 = (h - 0.5*(h[_n - 1] + h[_n + 1]))
gen num2 = (h - 0.5*(h[_n - 2] + h[_n + 2]))
gen num3 = (h - 0.5*(h[_n - 3] + h[_n + 3]))
gen denom = sqrt((3/2)*h)
gen SD1 = num1/denom
gen SD2 = num2/denom
gen SD3 = num3/denom


*NACE-SECTION R:
hist tot_revenue if (tot_revenue<=7000000 & tot_revenue>=3000000)  ///
 & tot_assets<20000000 & yr>2011          ///
 & (      ///
nace_11==90   | ///
nace_22==90   | ///
nace_11==91   | ///
nace_22==91   | ///
nace_11==92   | ///
nace_22==92   | ///
nace_11==93   | ///
nace_22==93   ///
),    ///
width(100000)          ///
xline(5000000)          ///
xlabel(3000000 "3" 4000000 "4" 5000000 "5" 6000000 "6" 7000000 "7") ///
xtitle("Total revenue, MNOK")         ///
frequency
```

*NACE-SECTION S:

```
hist tot_revenue if (tot_revenue<=7000000 & tot_revenue>=3000000)  ///
 & tot_assets<20000000 & yr>2011          ///
 & (      ///
 nace_11==94   | ///
 nace_22==94   | ///
 nace_11==95   | ///
 nace_22==95   | ///
 nace_11==96   | ///
 nace_22==96   ///
 ),    ///
 width(100000)          ///
 xline(5000000)           ///
 xlabel(3000000 "3" 4000000 "4" 5000000 "5" 6000000 "6" 7000000 "7") ///
 xtitle("Total revenue, MNOK")         ///
 frequency
```

*NACE-SECTION T:

```
hist tot_revenue if (tot_revenue<=7000000 & tot_revenue>=3000000)  ///
 & tot_assets<20000000 & yr>2011          ///
 & (      ///
 nace_11==97   | ///
 nace_22==97   ///
 ),    ///
 width(100000)          ///
 xline(5000000)          ///
 xlabel(3000000 "3" 4000000 "4" 5000000 "5" 6000000 "6" 7000000 "7") ///
 xtitle("Total revenue, MNOK")        ///
 frequency
```

*NACE-SECTION U:

```
hist tot_revenue if (tot_revenue<=7000000 & tot_revenue>=3000000)  ///
 & tot_assets<20000000 & yr>2011          ///
```

```
  & (       ///
 nace_11==99   | ///
 nace_22==99   ///
 ),     ///
 width(100000)          ///
 xline(5000000)          ///
 xlabel(3000000 "3" 4000000 "4" 5000000 "5" 6000000 "6" 7000000 "7") ///
 xtitle("Total revenue, MNOK")        ///
 frequency


********** BUNCHING PR. NACE-DIVISION **********
*NACE-DIVISION: 41
hist tot_revenue if (tot_revenue<=7000000 & tot_revenue>=3000000)  ///
 & tot_assets<20000000 & yr>2011         ///
 & (       ///
 nace_11==41   | ///
 nace_22==41   ///
 ),     ///
 width(100000)         ///
 xline(5000000)          ///
 xlabel(3000000 "3" 4000000 "4" 5000000 "5" 6000000 "6" 7000000 "7") ///
 xtitle("Total revenue, MNOK")        ///
 frequency

 *SD-test
 drop x h num1 num2 num3 denom SD1 SD2 SD3
 twoway__histogram_gen tot_revenue if (tot_revenue<=7000000    ///
 & tot_revenue>=3000000)         ///
 & tot_assets<20000000 & yr>2011         ///
 & (        ///
 nace_11==41    | ///
 nace_22==41    ///
```

```
),      ///
width(100000)          ///
frequency gen(h x)
gen num1 = (h - 0.5*(h[_n - 1] + h[_n + 1]))
gen num2 = (h - 0.5*(h[_n - 2] + h[_n + 2]))
gen num3 = (h - 0.5*(h[_n - 3] + h[_n + 3]))
gen denom = sqrt((3/2)*h)
gen SD1 = num1/denom
gen SD2 = num2/denom
gen SD3 = num3/denom


*NACE-DIVISION: 42
hist tot_revenue if (tot_revenue<=7000000 & tot_revenue>=3000000)  ///
 & tot_assets<20000000 & yr>2011          ///
 & (      ///
 nace_11==42    | ///
 nace_22==42     ///
 ),     ///
 width(100000)          ///
 xline(5000000)          ///
 xlabel(3000000 "3" 4000000 "4" 5000000 "5" 6000000 "6" 7000000 "7") ///
 xtitle("Total revenue, MNOK")          ///
 frequency


*SD-test
 drop x h num1 num2 num3 denom SD1 SD2 SD3
 twoway__histogram_gen tot_revenue if (tot_revenue<=7000000     ///
 & tot_revenue>=3000000)          ///
 & tot_assets<20000000 & yr>2011          ///
 & (          ///
 nace_11==42    | ///
 nace_22==42     ///
 ),     ///
```

```
width(100000)            ///
frequency gen(h x)
gen num1 = (h - 0.5*(h[_n - 1] + h[_n + 1]))
gen num2 = (h - 0.5*(h[_n - 2] + h[_n + 2]))
gen num3 = (h - 0.5*(h[_n - 3] + h[_n + 3]))
gen denom = sqrt((3/2)*h)
gen SD1 = num1/denom
gen SD2 = num2/denom
gen SD3 = num3/denom


*NACE-DIVISION: 43
hist tot_revenue if (tot_revenue<=7000000 & tot_revenue>=3000000)  ///
 & tot_assets<20000000 & yr>2011          ///
 & (      ///
 nace_11==43    | ///
 nace_22==43    ///
 ),      ///
 width(100000)           ///
 xline(5000000)           ///
 xlabel(3000000 "3" 4000000 "4" 5000000 "5" 6000000 "6" 7000000 "7") ///
 xtitle("Total revenue, MNOK")        ///
 frequency


*SD-test
drop x h num1 num2 num3 denom SD1 SD2 SD3
twoway__histogram_gen tot_revenue if (tot_revenue<=7000000     ///
 & tot_revenue>=3000000)          ///
 & tot_assets<20000000 & yr>2011         ///
 & (         ///
 nace_11==43    | ///
 nace_22==43    ///
 ),      ///
 width(100000)           ///
```

```
frequency gen(h x)
gen num1 = (h - 0.5*(h[_n - 1] + h[_n + 1]))
gen num2 = (h - 0.5*(h[_n - 2] + h[_n + 2]))
gen num3 = (h - 0.5*(h[_n - 3] + h[_n + 3]))
gen denom = sqrt((3/2)*h)
gen SD1 = num1/denom
gen SD2 = num2/denom
gen SD3 = num3/denom


*NACE-DIVISION: 45
hist tot_revenue if (tot_revenue<=7000000 & tot_revenue>=3000000)  ///
 & tot_assets<20000000 & yr>2011        ///
 & (      ///
 nace_11==45    | ///
 nace_22==45     ///
 ),    ///
 width(100000)           ///
 xline(5000000)          ///
 xlabel(3000000 "3" 4000000 "4" 5000000 "5" 6000000 "6" 7000000 "7") ///
 xtitle("Total revenue, MNOK")        ///
 frequency


*SD-test
 drop x h num1 num2 num3 denom SD1 SD2 SD3
 twoway__histogram_gen tot_revenue if (tot_revenue<=7000000     ///
 & tot_revenue>=3000000)          ///
 & tot_assets<20000000 & yr>2011         ///
 & (          ///
 nace_11==45    | ///
 nace_22==45     ///
 ),      ///
 width(100000)          ///
 frequency gen(h x)
```

```
gen num1 = (h - 0.5*(h[_n - 1] + h[_n + 1]))
gen num2 = (h - 0.5*(h[_n - 2] + h[_n + 2]))
gen num3 = (h - 0.5*(h[_n - 3] + h[_n + 3]))
gen denom = sqrt((3/2)*h)
gen SD1 = num1/denom
gen SD2 = num2/denom
gen SD3 = num3/denom


*NACE-DIVISION: 46
hist tot_revenue if (tot_revenue<=7000000 & tot_revenue>=3000000)  ///
 & tot_assets<20000000 & yr>2011          ///
 & (      ///
 nace_11==46    | ///
 nace_22==46     ///
 ),     ///
 width(100000)          ///
 xline(5000000)          ///
 xlabel(3000000 "3" 4000000 "4" 5000000 "5" 6000000 "6" 7000000 "7") ///
 xtitle("Total revenue, MNOK")        ///
 frequency


*SD-test
 drop x h num1 num2 num3 denom SD1 SD2 SD3
 twoway__histogram_gen tot_revenue if (tot_revenue<=7000000     ///
 & tot_revenue>=3000000)           ///
 & tot_assets<20000000 & yr>2011         ///
 & (          ///
 nace_11==46    | ///
 nace_22==46     ///
 ),     ///
 width(100000)          ///
 frequency gen(h x)
 gen num1 = (h - 0.5*(h[_n - 1] + h[_n + 1]))
```

```
gen num2 = (h - 0.5*(h[_n - 2] + h[_n + 2]))

gen num3 = (h - 0.5*(h[_n - 3] + h[_n + 3]))

gen denom = sqrt((3/2)*h)

gen SD1 = num1/denom

gen SD2 = num2/denom

gen SD3 = num3/denom


*NACE-DIVISION: 47
hist tot_revenue if (tot_revenue<=7000000 & tot_revenue>=3000000)  ///
 & tot_assets<20000000 & yr>2011          ///
 & (      ///
 nace_11==47    | ///
 nace_22==47     ///
 ),    ///
 width(100000)         ///
 xline(5000000)         ///
 xlabel(3000000 "3" 4000000 "4" 5000000 "5" 6000000 "6" 7000000 "7") ///
 xtitle("Total revenue, MNOK")       ///
 frequency


*SD-test
 drop x h num1 num2 num3 denom SD1 SD2 SD3
 twoway__histogram_gen tot_revenue if (tot_revenue<=7000000     ///
 & tot_revenue>=3000000)          ///
 & tot_assets<20000000 & yr>2011         ///
 & (          ///
 nace_11==47    | ///
 nace_22==47     ///
 ),    ///
 width(100000)         ///
 frequency gen(h x)
 gen num1 = (h - 0.5*(h[_n - 1] + h[_n + 1]))
 gen num2 = (h - 0.5*(h[_n - 2] + h[_n + 2]))
```

```
gen num3 = (h - 0.5*(h[_n - 3] + h[_n + 3]))
gen denom = sqrt((3/2)*h)
gen SD1 = num1/denom
gen SD2 = num2/denom
gen SD3 = num3/denom


*NACE-DIVISION: 49
hist tot_revenue if (tot_revenue<=7000000 & tot_revenue>=3000000)  ///
& tot_assets<20000000 & yr>2011          ///
& (        ///
nace_11==49   | ///
nace_22==49     ///
),     ///
width(100000)          ///
xline(5000000)          ///
xlabel(3000000 "3" 4000000 "4" 5000000 "5" 6000000 "6" 7000000 "7") ///
xtitle("Total revenue, MNOK")        ///
frequency


*SD-test
drop x h num1 num2 num3 denom SD1 SD2 SD3
twoway__histogram_gen tot_revenue if (tot_revenue<=7000000     ///
& tot_revenue>=3000000)          ///
& tot_assets<20000000 & yr>2011          ///
& (          ///
nace_11==49   | ///
nace_22==49     ///
),     ///
width(100000)          ///
frequency gen(h x)
gen num1 = (h - 0.5*(h[_n - 1] + h[_n + 1]))
gen num2 = (h - 0.5*(h[_n - 2] + h[_n + 2]))
gen num3 = (h - 0.5*(h[_n - 3] + h[_n + 3]))
```

```
gen denom = sqrt((3/2)*h)

 gen SD1 = num1/denom

 gen SD2 = num2/denom

 gen SD3 = num3/denom


*NACE-DIVISION: 50
hist tot_revenue if (tot_revenue<=7000000 & tot_revenue>=3000000)  ///
 & tot_assets<20000000 & yr>2011          ///
 & (       ///
 nace_11==50    | ///
 nace_22==50     ///
 ),    ///
 width(100000)            ///
 xline(5000000)            ///
 xlabel(3000000 "3" 4000000 "4" 5000000 "5" 6000000 "6" 7000000 "7") ///
 xtitle("Total revenue, MNOK")       ///
 frequency


*SD-test
 drop x h num1 num2 num3 denom SD1 SD2 SD3
 twoway__histogram_gen tot_revenue if (tot_revenue<=7000000     ///
 & tot_revenue>=3000000)          ///
 & tot_assets<20000000 & yr>2011          ///
 & (          ///
 nace_11==50    | ///
 nace_22==50     ///
 ),     ///
 width(100000)            ///
 frequency gen(h x)
 gen num1 = (h - 0.5*(h[_n - 1] + h[_n + 1]))
 gen num2 = (h - 0.5*(h[_n - 2] + h[_n + 2]))
 gen num3 = (h - 0.5*(h[_n - 3] + h[_n + 3]))
 gen denom = sqrt((3/2)*h)
```

105

```
gen SD1 = num1/denom
gen SD2 = num2/denom
gen SD3 = num3/denom


*NACE-DIVISION: 51
hist tot_revenue if (tot_revenue<=7000000 & tot_revenue>=3000000)  ///
 & tot_assets<20000000 & yr>2011        ///
 & (      ///
 nace_11==51    | ///
 nace_22==51     ///
 ),      ///
 width(100000)          ///
 xline(5000000)          ///
 xlabel(3000000 "3" 4000000 "4" 5000000 "5" 6000000 "6" 7000000 "7") ///
 xtitle("Total revenue, MNOK")        ///
 frequency


*SD-test
drop x h num1 num2 num3 denom SD1 SD2 SD3
twoway__histogram_gen tot_revenue if (tot_revenue<=7000000      ///
 & tot_revenue>=3000000)         ///
 & tot_assets<20000000 & yr>2011         ///
 & (        ///
 nace_11==51    | ///
 nace_22==51     ///
 ),      ///
 width(100000)          ///
 frequency gen(h x)
gen num1 = (h - 0.5*(h[_n - 1] + h[_n + 1]))
gen num2 = (h - 0.5*(h[_n - 2] + h[_n + 2]))
gen num3 = (h - 0.5*(h[_n - 3] + h[_n + 3]))
gen denom = sqrt((3/2)*h)
gen SD1 = num1/denom
```

```
 gen SD2 = num2/denom
 gen SD3 = num3/denom


*NACE-DIVISION: 52
hist tot_revenue if (tot_revenue<=7000000 & tot_revenue>=3000000)  ///
 & tot_assets<20000000 & yr>2011          ///
 & (      ///
 nace_11==52    | ///
 nace_22==52     ///
 ),    ///
 width(100000)          ///
 xline(5000000)          ///
 xlabel(3000000 "3" 4000000 "4" 5000000 "5" 6000000 "6" 7000000 "7") ///
 xtitle("Total revenue, MNOK")        ///
 frequency


*SD-test
 drop x h num1 num2 num3 denom SD1 SD2 SD3
 twoway__histogram_gen tot_revenue if (tot_revenue<=7000000     ///
 & tot_revenue>=3000000)         ///
 & tot_assets<20000000 & yr>2011          ///
 & (         ///
 nace_11==52    | ///
 nace_22==52     ///
 ),     ///
 width(100000)          ///
 frequency gen(h x)
 gen num1 = (h - 0.5*(h[_n - 1] + h[_n + 1]))
 gen num2 = (h - 0.5*(h[_n - 2] + h[_n + 2]))
 gen num3 = (h - 0.5*(h[_n - 3] + h[_n + 3]))
 gen denom = sqrt((3/2)*h)
 gen SD1 = num1/denom
 gen SD2 = num2/denom
```

```
gen SD3 = num3/denom


*NACE-DIVISION: 53
hist tot_revenue if (tot_revenue<=7000000 & tot_revenue>=3000000)  ///
 & tot_assets<20000000 & yr>2011          ///
 & (      ///
 nace_11==53    | ///
 nace_22==53     ///
 ),     ///
 width(100000)          ///
 xline(5000000)          ///
 xlabel(3000000 "3" 4000000 "4" 5000000 "5" 6000000 "6" 7000000 "7") ///
 xtitle("Total revenue, MNOK")       ///
 frequency


*SD-test
 drop x h num1 num2 num3 denom SD1 SD2 SD3
 twoway__histogram_gen tot_revenue if (tot_revenue<=7000000     ///
 & tot_revenue>=3000000)          ///
 & tot_assets<20000000 & yr>2011          ///
 & (        ///
 nace_11==53    | ///
 nace_22==53     ///
 ),      ///
 width(100000)          ///
 frequency gen(h x)
 gen num1 = (h - 0.5*(h[_n - 1] + h[_n + 1]))
 gen num2 = (h - 0.5*(h[_n - 2] + h[_n + 2]))
 gen num3 = (h - 0.5*(h[_n - 3] + h[_n + 3]))
 gen denom = sqrt((3/2)*h)
 gen SD1 = num1/denom
 gen SD2 = num2/denom
 gen SD3 = num3/denom
```

```
*NACE-DIVISION: 55
hist tot_revenue if (tot_revenue<=7000000 & tot_revenue>=3000000)  ///
 & tot_assets<20000000 & yr>2011         ///
 & (     ///
 nace_11==55    | ///
 nace_22==55    ///
 ),     ///
 width(100000)          ///
 xline(5000000)          ///
 xlabel(3000000 "3" 4000000 "4" 5000000 "5" 6000000 "6" 7000000 "7") ///
 xtitle("Total revenue, MNOK")       ///
 frequency


*SD-test
 drop x h num1 num2 num3 denom SD1 SD2 SD3
 twoway__histogram_gen tot_revenue if (tot_revenue<=7000000     ///
 & tot_revenue>=3000000)         ///
 & tot_assets<20000000 & yr>2011        ///
 & (        ///
 nace_11==55    | ///
 nace_22==55     ///
 ),      ///
 width(100000)          ///
 frequency gen(h x)
 gen num1 = (h - 0.5*(h[_n - 1] + h[_n + 1]))
 gen num2 = (h - 0.5*(h[_n - 2] + h[_n + 2]))
 gen num3 = (h - 0.5*(h[_n - 3] + h[_n + 3]))
 gen denom = sqrt((3/2)*h)
 gen SD1 = num1/denom
 gen SD2 = num2/denom
 gen SD3 = num3/denom
```

```
*NACE-DIVISION: 56
hist tot_revenue if (tot_revenue<=7000000 & tot_revenue>=3000000)  ///
 & tot_assets<20000000 & yr>2011          ///
 & (       ///
 nace_11==56    | ///
 nace_22==56     ///
 ),      ///
 width(100000)        ///
 xline(5000000)        ///
 xlabel(3000000 "3" 4000000 "4" 5000000 "5" 6000000 "6" 7000000 "7") ///
 xtitle("Total revenue, MNOK")         ///
 frequency


*SD-test
 drop x h num1 num2 num3 denom SD1 SD2 SD3
 twoway__histogram_gen tot_revenue if (tot_revenue<=7000000     ///
 & tot_revenue>=3000000)          ///
 & tot_assets<20000000 & yr>2011         ///
 & (          ///
 nace_11==56    | ///
 nace_22==56     ///
 ),     ///
 width(100000)           ///
 frequency gen(h x)
 gen num1 = (h - 0.5*(h[_n - 1] + h[_n + 1]))
 gen num2 = (h - 0.5*(h[_n - 2] + h[_n + 2]))
 gen num3 = (h - 0.5*(h[_n - 3] + h[_n + 3]))
 gen denom = sqrt((3/2)*h)
 gen SD1 = num1/denom
 gen SD2 = num2/denom
 gen SD3 = num3/denom
```