

Valuation Fundamentals in Emerging Markets

Mustafa Ablay

Abstract

This thesis looks into the performance of equity analysts in carrying out discounted cash flow (DCF) valuations in an emerging market setting, in respect to Kazakhstan and other frontier markets. A new data set of 106 analyst valuation models (2010-2025) is assembled to examine the subjective discount rates and growth expectations contained in these DCFs. This work record assumes discount rates by analysts are high (mean 14.6%) but with significant variation with firms (7-19%), and that the terminal growth rates are low (median 2.0) with many being zero or low even when there is a high level of inflation. The terminal value is added on an average basis of 56 percent of the total DCF value and that despite the emerging market setting, most of the value is represented by long-term projections. This thesis discovers by cross sectional and time series analysis that both growth expectations and discount rates are significant explanatory variables of valuation differences. In line with the most recent results of Décaire and Graham (2024), the thesis demonstrates that differences in subjective discount rates arise mainly due to differences in the global risk free rate and individual company β s and not local country risk or inflation. This filtering out of volatile local market conditions by analysts with constant USD-denominated risk-free rates, and equity risk premia, seems to be consistent with Bayesian updating under imperfect information. The growth assumptions on the terminals are weakly related to the real GDP growth projections and they demonstrate no relation to the current inflation, implying that analysts anchor long-run growth to real fundamentals as opposed to nominal models. This paper also discuss the internal consistency of these valuation inputs: the portion of the total value attributable to near-term cash flows versus that attributable to the terminal period, the sensitivity of discount rates to the timing of future predictions of cash flows (it is nonexistent), and whether quantitative adjustments to extreme assumptions (winsorizing) has a significant effect on valuation (it does not). In general, results present an in-depth image of practitioner use of DCF in an evolving market environment - of global valuation fundamentals versus local risk - and the paper comment on the implication of the results to investment strategy and academic valuation models when operating in high-volatility markets.

1. Introduction

The analysts who price the firms in frontier emerging markets are presented with a specific challenge: a model to price volatile macroeconomic conditions with double-digit inflation rates, exchange-rate fluctuations, shallow capital markets, etc., requires few DCF parameters including discount rates and long-run growth. The standard theory assumes that these inputs would be systematically distributed with regards to macro conditions. However, anecdotal experience indicates that, in most cases, analysts base their analysis on world benchmarks, impose informal floors on discount rates, and use low terminal growth when local nominal conditions are either very extreme. This behavior is systematic, rational or consistent with the findings of the developed markets and whether this is so or not is a question that is open to discussion.

Evidence from large samples has recently demonstrated the joint role of subjective discount rates and growth expectations in determining valuation outcomes in developed markets, and that analysts continuously update both inputs in a manner that is consistent with rational (Bayesian) learning. Nonetheless, the question is whether such trends are applicable in the emerging markets where the markets are characterized by a significant disparity in the market inefficiencies and risk perceptions. As an illustration, a lot of emerging market values have been experiencing very high local interest rates and inflation which makes it questionable whether the local or global risk-free rate is applied by the analysts. Also, as indicated in earlier research and practitioner surveys, terminal value might represent an even higher proportion of total value in new markets, because forecast horizons are shorter and there is more uncertainty about the future cash flows within a relatively short timeframe (Bruner et al., 2002). This study offers an insight on how practitioners model the DCF fundamentals to suit such a situation dataset of analyst DCF valuations in Kazakhstan, a frontier market with periodic crises and structural reforms.

This thesis has three contributions to the literature of valuation.

To start with, it presents a novel sample of 106 analyst DCF valuation models of firms in Kazakhstan and similar frontier markets. This is, to the best of available evidence, the first structured collection of analyst valuation input to be applied to Central Asia, which is a geographic and methodological gap in literature.

Second, it records the process of establishment of subjective discount rates and terminal growth by analysts in a macroeconomically unstable environment. There are indications that discount rates are concentrated around 14-16 percent even though they can vary significantly in local inflation and bond rates indicating anchoring to USD standards and informal WACC floors, which has not been observed in frontier markets.

Third, this work conducts tests on how the findings of Décaire and Graham (2024) can be applied to frontier markets. It also discovered that terminal value almost explains more than half of intrinsic value, that terminal growth is very insensitive to inflation, and that discount rates partly pass through to variations in the global risk-free rate and β . These findings imply that analysts reason in real terms where there is a high inflation environment and that macro volatility is filtered in a way that is compatible with Bayesian rationalization.

Altogether, such contributions provide the first empirical characterization of the fundamentals of valuation in a frontier emerging market and give new evidence about how standard valuation frameworks guide practitioners in adapting during macroeconomic uncertainty.

2. Literature Review

The practice of fundamental valuation has been a subject of discussion in both the literature on investment and corporate finance. The literature (e.g. Damodaran, 2012) discusses how analysts ought to estimate the discount rates and growth to use in DCF models, yet the empirical studies indicate that there is a great variety, and biases exist in these inputs. Initial research by Kaplan and Ruback (1995) and Bruner, Eades, Harris, and Higgins (1998) revealed that on average DCF valuations may model market prices but that the terminal value assumptions used tend to delegate most of the difference in outcomes. As a matter of fact, the practitioners understand that much of the value of a DCF arises during the terminal period - which has been quoted as 50-80 percent under common assumptions - and this casts doubt on the subjectivity of making long-run growth and exit multiples. As the recent studies by Décaire and Graham (2024) affirm, the terminal value usually covers approximately 75 percent of the total valuation in the model of the analysts, it is important to know how terminal growth rates are determined.

Several studies have been conducted regarding the determination of discount rates (cost of equity or WACC) by analysts. In developed economies, the capital asset pricing model (CAPM) is

typically implied with a risk-free interest rate (typically a long-term government bond yield) and an equity risk premium (ERP) multiplied by the β of a stock. Nevertheless, subjective modulations are common - analysts may add size premiums, idiosyncratic risk factors or, simply, round-number discount rates, which do not follow CAPM. Analyst forecasts have implied cost of equity which has significant cross-sectional variations that cannot be explained by CAPM β 's alone indicating analysts use private information or non- β risks. They discover that subjective discount rates determined by analysts generally adhere to CAPM intuition - that they vary over time primarily due to the risk-free rate and β 's - but also that the returns required by analysts do explain future stock returns (i.e. do contain real information about risk). It means that the discount rates of analysts do not consist of noise alone; they contain the expectation of risk to be consistent with future results (a steep subjective security market line).

The further expansion of these ideas to the new markets brings in more complexity. Damodaran (2015) and Pereiro (2006) recommend that in the emerging markets, valuations are usually characterized by the addition of a country risk premium to the global equity risk premium to reflect the political and economic risks. One of the methods suggests that the yield on USD-denominated state bonds should be used as a proxy to country risk - in effect being a boost to the discount rate of companies based in countries with greater country risk. Another way is however to manipulate cash flows (e.g. scenario probabilities) without changing the discount rates and ensure that they are pegged to world standards. It is still being debated whether to add country risk to the discount rate, or cash flows; analysts differ. An example would be to have country risk implicitly modeled as increased volatility of cash flows or in the equity risk premium using a global risk-free rate (such as U.S. 10-year Treasury yield) to all values. The data will reveal the popular approach in Kazakhstan.

Projections on the growth by analysts are also worth consideration. In earlier studies (Dechow and Sloan, 1997; Richardson et al., 2006), analysts have been found to be optimistic when making short-term forecasts of earnings, but the decision they make regarding long-term (terminal) growth remains less reported. According to Décaire and Graham (2024), the new evidence confirms that the growth estimates of terminals that are made by analysts are tracked with the real GDP growth and have little relation to inflation. This curious finding indicates that the analysts are thinking about terminal growth on real terms - e.g. in a low-inflation developed

economy, a terminal growth of about 2 percent may be close to real GDP growth. In emerging markets where inflation is higher, when analysts anchor on the same parameter of real GDP, their terminal growth assumptions would be very low in comparison to the nominal local rates of economic growth. This may represent the belief that in the long run, depreciation of local currency cancels inflation (when measured in a hard currency) or it may be conservative bias. The specific test that The thesis will be applying in the thesis will be that the terminal growth rates applied in Kazakhstan are more related to real economic growth or inflation.

And lastly it appeals to literature relating to behavioral and Bayesian updating by analysts. Décaire and Graham (2024) explain the patterns of analysts by rational filtering: when given noisy information on fundamentals (e.g. volatile bond yields), analysts change their estimates slowly, which is in line with Bayesian updating in the face of uncertainty. To add to this debate, the thesis will look at the way analysts in the sample had to adjust their discount rates to large external shocks (e.g. interest rate spikes or crises). When they partially bend the risk-free element of WACC in a response to the local rates soaring, then this can be a sign that they think the shock is non-persistent noise - a Bayesian implication. This work will observe that Kazakhstan centric analysts have not wholly integrated surging domestic bond yields into increased discount rates, effectively ironing volatility. This is in line with rational filtering (local rate spikes are temporary) and may be limited by conventions (e.g. always using a U.S.-dollar risk-free benchmark).

2.1 Research Gap

Despite the contribution of the previous research, which gives substantial evidence on the way analysts make their discount rates and long-run growth expectations in the developed markets, there is nearly no credible systematic work on the fundamentals of the DCF in the frontier emerging markets. Available literature is mainly based on U.S. or Western European analyst information, in which macroeconomic factors are fairly stable and capital markets are deep. Subsequently, there is a question of whether the patterns observed by Décaire and Graham (2024) including the conservative terminal growth assumptions, incomplete Bayesian updating of discount rates, and dominance of terminal value are also present in high-volatility environments where the inflation, currency depreciation and sovereign risk are substantial.

Moreover, most of the valuation literature on emerging markets comment on methodological adaptations (e.g., country risk premia, local vs. global risk-free rates), but has no empirical data on what analysts actually practice in practice when making cash flows forecasts or determining WACC and terminal growth. To the best of the available literature, there is no study published that has studied hand-collected analyst valuation models in Central Asia or other frontier economies.

This gap is filled by this thesis through the compilation of the first dataset of analyst DCF models of large Kazakhstani firms (2010-2025), empirical accounting of the subjective discount rates, growth assumptions, and the mechanics of valuation as applied in practice. This permits us to evaluate the existence of regularities in the developed markets in a frontier environment, and it serves to uncover market-specific behavior, including hard floors in the WACC or USD anchoring, that is not apparent in preexisting literature.

Overall, this thesis is based on the previous literature as it offers a concentrated discussion of valuation inputs in a developing market. It is a test of whether the patterns that have been observed in the developed markets (the significance of discount rates and expansion in valuation, CAPM drivers of discount rates, GDP connection of terminal expansion, etc.) are valid, and the emerging market environment modifies those dynamics using country risk and macro volatility.

3. Institutional Background

Kazakhstan is an exemplary frontier emerging economy which has unique institutional factors that contribute to valuation practice. The stock market of the country (Kazakhstan Stock Exchange, KASE) is not very young and small. The local stock exchange is not very active with only several dozen companies listed and most stocks have low liquidity with the exception of the blue-chip companies (e.g. national companies in banking, natural resources, telecom). Most Kazakhstani companies are also listed or depository receipts outside the country (in London or Hong Kong) and, consequently, analysts in these companies are frequently reading an international investor. The pool of domestic investors is small, as well as the coverage of analysts - usually on the biggest of them. Consequently, every available analyst report is an opportunity to gain some information on assumptions being made about risk and growth.

The economy and financial markets of Kazakhstan have been subject to a lot of volatility within the last 10 years. As a commodity-based economy (oil, metals), Kazakhstan experienced boom and bust periods due to commodity price fluctuations. As an example, the oil price meltdown of 2014-2015 resulted in a dramatic currency devaluation (the Kazakhstani tenge dropped more than 50 percent in 2015) and an inflation spurt. This increased interest rates when the National Bank used a free-float exchange rate and set policy rates higher than in fighting inflation. The direct effect of those events on valuation parameters is that the 10-year government bond yield increased to the middle of the teens (it even hit 16-17% at one point and another). Conversely, the stable periods of time (2010-2013) had a moderate inflation (5-7) and bond yields of about 7-8. Inflation rates in Kazakhstan returned to the 2020-2021, with inflation surging over 8 per cent and then the surge of the global inflation in 2022 (inflation >14%).

This is an issue of valuation dilemma due to such macro instability. One would simply put the local risk-free rate, say of 15%, into a CAPM model - but so would result in very high discount rates (possibly 20% and above), which would underprice companies in the event that the high yields will be reduced. In fact, a large number of practitioners in emerging economies prefer to appreciate a stable currency (usually USD) to prevent the local rates of inflation. In the example of Kazakhstan, most analytic reports are written in U.S. dollars (when reporting on dual-listed companies), or simply make an assumption of a long-term U.S. Dollar to Kazakhstan exchange rate trend in order to isolate real growth and inflation. The sample is approximated to be 4% risk-free rates applied by analysts, and this is closely similar to the U.S. Treasury yield in the 2010s and mid-2020s - highly indicating a U.S. dollar risk-free rate is being applied by the analysts. This is a very important institutional aspect: although Kazakhstan has high local yields, the USD risk-free rate is frequently used as the input and the country risk is then calculated using the equity risk premium or cash flow haircuts instead of directly using the risky KZT yield.

The capital market regulations in Kazakhstan have stimulated increased transparency and research in the past few years. In 2018, the so-called Astana International Financial Centre (AIFC) was opened, the purpose of which is to facilitate foreign investments; new structures that are consistent with international standards have been introduced. Certain significant privatizations (e.g. Kazatomprom IPO in 2018, Halyk Bank SPO, KazMunayGas IPO in 2022) created a situation where they are more covered by foreign banks and local brokers. These

analysts, like those in other emerging markets (Russia, Eastern Europe), introduce methodologies akin to those of other emerging markets, such as frequently a two stage DCF, explicit forecast at most 5 years and terminal value typically based on a perpetuity growth model. In such markets, a fair degree of conservativeness in terminal growth (typically pegged to long-term global growth or inflation differentials) and either blended equity risk premium (global + country spread) or scenario-based changes is an institutional norm.

As an example, the country risk in Kazakhstan is rated moderate-high (sovereign credit rating BBB- to BB+ in the period). Damodaran (2012) recommends that the default spread of the country should be added to the equity risk premium. In recent years, the USD-based bonds of Kazakhstan have a yield of approximately 1.5-2% over the U.S. Treasuries and this may be translated as an extra 2% in the equity risk premium of Kazakh stocks. As a matter of fact, there are lots of analysts applying an equity risk premium of 7.5-10, which is more than the 5-6 that most people tend to assume in the U.S. This implies that they are bringing country risk to the ERP.

The other institutional dimension is the quality of financial reporting among the emerging markets. The blue-chip companies of Kazakhstan are typically under IFRS and with Big-4 audits; however, concerns of governance (state ownership, related-party transaction) might make the analysts add more risk premiums or use a more conservative forecast. On the example of an oil company, which is majority-owned by the state, analysts may forecast reduced margin growth or an increased risk element of possible quasi-fiscal obligations (e.g. providing the domestic market at reduced costs). Such considerations can be effectively considered as the adjustments to the cash flow as opposed to discount rate. The thesis has a dataset of final WACC and growth assumptions of the analysts but the story in their reports usually gives a reason for these qualitative habits. The institutional implication is that the new market valuations tend to be composed of overlays - some based on the discount rate (greater ERP, β adjustments), some on the cash flows (conservative forecasts, scenario probabilities).

In a nutshell, the environment in Kazakhstan compels analysts to deal with high inflation/interest rate fluctuations, currency devaluations and government interference. The current trend seems to be decoupling to local noise: value in hard currency terms, use international standard in risk free rate and local risk through ERP or β . This enables the analysts to come up with more stable and

comparable valuations across the borders as opposed to the seesaw valuations due to every macro spike. This thesis shall find the manifestation of such a method in the results where the subjective discount rates employed by analysts are confined within a narrow band of between +12-17% despite the fact that the local bond yields are between 7 and 16%. This regularity implies the institutional agreement of managing emerging market risks in DCF models that, actually, associates with rationalizing the temporary local environment (as postulated by Décaire and Graham, 2024).

4. Data and Methodology

The first step would be to create a sample of analyst valuation. Then gather equity research reports of 2010 to 2025 of large Kazakhstani firms. The next step is to use investment bank and local brokerage firm reports, in PDF format, which will contain a detailed output of the DCF model. Based on every report, the thesis extract the important valuation inputs manually, including the discount rate (usually the Weighted Average Cost of Capital, WACC, or cost of equity), the explicit forecast horizon (number of years of estimated cash flows before the terminal value), the annual forecasts of cash flows and the terminal growth rate (G_T) applied in the terminal value calculation. As a result 106 different valuation cases of 15 firms were obtained.

The work concentrates on the DCF models employing the two-stage traditional two-stage traditional setup: explicit forecast period of N years, and a terminal value of a growing perpetuity. The common example in the sample is to assume free cash flow (to the firm or the equity) of $N \sim 5-7$ years and assume a terminal growth rate of G_T to obtain the terminal value at the N th year. Discount rate is WACC (when the flow is to the firm) or cost of equity (when the flow to equity - most reports in the sample on that basis report WACC and value the firm and debt will produce equity value). To be consistent, this thesis takes all reported values of discount rates as the required rate of return by the analysts on the enterprise, and all values of terminal growth rates as (nominal) growth in the nominal free cash flow after the horizon.

At the end the summary of these data to examine distributional characteristics and correlations. Section 4.1 (Table 1) has presented summary statistics (means, medians, ranges) of the key inputs. Also computation of the amount of valuation of each valuation terminal value share - i.e.

the proportion of the total DCF amount attributable to the terminal period. This is also calculated by reestablishing the model inputs into the implied enterprise value by discounting the projected cash flows and terminal value and then dividing the ratio of the PV of the terminal value by the cumulative PV. To ensure strength, the thesis checks that the values of the total it construct are consistent with the reported target prices (they are, to small rounding errors, hence the extraction consistency). In some instances the reports did not clearly mention G_T (e.g. where an analyst has used an exit multiple or n.a. in text); The thesis do not include these in analyses of G_T or terminal share (9 cases).

To test the determinants of the selected discount rates, the thesis employs cross-sectional comparisons as well as the time-series analysis. The mismatch each date of valuation with macroeconomic data of the day: the 10-year U.S. Treasury yield (as a proxy of world risk-free rate), the 10-year yield of the government bond market in Kazakhstan (as a proxy of local risk-free), and general market measures of equity risk. The thesis also pulled out reports where the thesis could include the elements of WACC; risk-free rate assumption, equity risk premium and β . Then it carried out a regression analysis (Section 6.2) of the type:

$$WACC_{i,t} = \alpha + \beta_1(Risk - free)_t + \beta_2(Beta_i \times ERP_t) + \varepsilon_{i,t}$$

where i, t indexes firm and time. This is inspired by the CAPM formula; if analysts strictly followed CAPM, we'd expect $\beta_1 \approx 1$, $\beta_2 \approx 1$, $\alpha \approx 0$, (no constant bias). Due to the small sample (a particular one in terms of explicit risk-free and ERP components at the disposal), this regression can be viewed as more of an illustrative one.

The thesis also tests the association between growth rate of terminals and the macro variables of the analysts. The thesis takes annual real GDP growth and inflation of Kazakhstan corresponding with the year of report of each valuation. Because analyst terminal growth is a long-run assumption, this paper will hypothesize it to be associated with the average expected real GDP growth (e.g. an analyst who expected the economy to grow at 4% on long-run average projected that they would set the G_T at 4% in real terms, or 4% + inflation). But when analysts are operating with hard-currency (USD) models, their G_T may represent growth rates in the world (or the developed markets) in general (say 2-3) notwithstanding local GDP. Using the correlation

with the growth and inflation of Kazakhstan, the thesis can guess their anchor. These correlations are as shown in Section 6.3 (Table 6).

In order to know the contribution of each of the valuations of the various periods, it computes the present value contribution of blocks of cash flows. In particular, the thesis determines: - Block 1: the present value of cash flows of years 1-3 of the forecast. - Block 2: the present value of cash flows between years 4-N (rest of the explicit period). - Block 3: present value of the terminal value.

These are calculated using reported WACC, per valuation. (Where valuation $N < 3$, block the work covers to N). Then the thesis obtains the portion of the total value of each block. Section 6.1 presents summary statistics and a pictorial chart of these block shares in terms of their horizon length. This block analysis assists in answering questions such as: do long-horizon forecasts in fact minimize dependence on terminal value (as would otherwise happen)? What is the importance of the initial years of the cash flow to the overall evaluation (in particular in emerging markets where the cash flows in the short-term may be low or inconsistent)?

Potential limitations of the data should be mentioned. The sample, though extensive in relation to covering the active coverage in Kazakhstan, is small. There are certain companies (such as Kaspi.kz, a fintech company) that have more than one report in recent years, which may make their impact too powerful, and some that are only present once. To reduce this, in cross-sectional studies the theses have occasionally averaged several valuations of a firm to equalize the firms (especially in the ERP vs β chart in Section 6.4). In addition, all the data are reported by the analysts - The thesis suppose that they are internally consistent. The thesis also identified one or two minor anomalies (e.g. a report of WACC = 14,4% which should have been WACC = 14.4% which the thesis had converted to 0.144 in the decimal point) and corrected those in hand.

To conclude, the approach will be a mixture of descriptive statistics (to answer the question what is the typical WACC or G_T in this market?), comparative statistics (to relate inputs to macro or cross-firm to answer the question why do these inputs differ?), and reconstruction analysis (e.g. to break down the DCF value into components to answer the question how these interact with cash flow timing and discount rates?). Through the triangulation of these methods, the experts expect to develop an in-depth picture of the basics of valuation in an emerging market.

5. Empirical Results

5.1 DCF Input Descriptive Statistics

The starting point is to summarize the most important DCF inputs in the sample of the emerging market. Table 1 shows the descriptive statistics of the discount rates and terminal growth rates that the analysts employ and the corresponding contributions of the terminal value to the total enterprise value. To provide the comparison, the thesis also presents average values indicated in Décaire and Graham (2024) in a developed-market setting.

Table 1. DCF Valuation Inputs Summary

(Terminal Value Share is the percentage of total DCF value coming from the terminal period. Standard Deviation in parentheses.)

Metric	Mean	Median	Min - Max
Discount Rate (WACC)	14.6% ($\pm 2.7\%$)	15.5%	7.0% - 19.1%
Risk-Free Rate (implied)	$\sim 4.3\%$	-	- (used USD 10y $\sim 4\%$)
Equity Risk Premium (implied)	$\sim 8.0\%$	-	- (varies 7%-10%)
Levered β (implied)	~ 0.85	-	- (0.7-1.3 range)
Terminal Growth Rate (G_T)	1.99% ($\pm 1.04\%$)	2.0%	0% - 4.1%
Terminal Value Share	56.0% ($\pm 19.1\%$)	55.2%	0% - 82.9%

Source: Analyst reports dataset (106 valuations). Developed market norms for WACC $\sim 8\text{-}10\%$, $G_T \sim 2\text{-}4\%$, terminal share $\sim 70\text{-}80\%$ for comparison.

A number of observations come up. Discount rates are extremely high in Kazakhstan: the average WACC is 14.6%, and the median discount range of 50% is between 13.4% and 16.5%. Such rates are approximately two times higher than the average rates in the U.S. or European markets (where WACC usually ranges between 7-10 percent). The high discount rates imply the extra risk and expectations of inflation in an emerging market. It is, however, noteworthy that the risk-free element of these WACCs is significantly small as compared to the yields on local government debt. The implied risk-free rate is on average of 4.3 per cent which aligns with U.S. Treasury yields in the sample run - which validates that analysts anchor on a developed-market

risk-free rate (USD or EUR) and not on the Kazakhstani 10-year bond (which averaged 8-12% in the 2010s and 2020s and hit 15 per cent in high-inflation periods). The implied equity risk premium (ERP) of approximately 8 percent, exceeds common assumptions of the U.S. ERP (5-6 percent), implying a country risk premium of 2-3 percent on top of other global equity risk expectations. ERP values were explicitly mentioned in different reports in the sample of 7.4%, 7.9%, 9.7%, 10.4% which were entirely consistent with this interpretation. The stated levered β s are centered around 0.8-1.0 in case of large stable companies, and up to 1.25 in more risky cases (eg. Air Astana, an airline).

The terminal growth rates (G_T) of analysts are very low when considered in absolute terms. The median stands at 2.0%, and the entire 60 per cent of the valuations presuppose terminal growth of between 2 and 2.5. The rest are predominantly at 0% or 3; 3 cases are above 3 and the highest is at 4.1. Given that inflation in Kazakhstan used to be higher than 5 percent, a terminal growth of 2 percent means a very conservative nominal growth or assuming long-run real growth of about zero (assuming 2 percent is supposed to be counterbalanced by inflation). Indeed, 15 (15/100) of the sample, which is 15 valuations, have $G_T G_T = 0\%$ - in effect, no growth at all in the future, which can either indicate a belief of stagnation or indicate a wish to be extremely careful with unstable companies. Conversely, none of them assumed G_T over 5% which one would consider in case one was bound to local inflation remains plus the real growth. This validates the fact that, as with their developed-market counterparts, analysts base terminal growth on world long-term growth standards (low single digits) as opposed to local inflation. This is measured in Section 6.3: in the sample, G_T is significantly associated with real GDP growth and virtually not with inflation.

What is the implication of these inputs on a share of value coming out of the terminal period? The mean and median of the terminal value share are 56% and 55% respectively. This value falls short of the 75% commonly quoted to the U. S. markets but remains above half of overall value on average. The spread is extensive: terminals in some instances contribute merely a fraction of value, (i.e. <30-40 percent), in others value is above 70 percent. The maximum and minimum terminal share values (40% and less) occur in a small number of cases where the initial cash flows (year 1) were very large (e.g. a mining project with huge cash flows in the first year of

operation) or where the analyst truncated the forecast and did not have a standard perpetuity (two cases of troubled firms effectively valued at the asset basis, giving terminal value 0%). On the other hand, the largest terminal values (75-83 percent) are given to companies with small cash flow in the near-term and with a long growth trajectory (e.g. a fast-growing fintech where most of the value is estimated after 7 years of growth). Excluding extremes, the central tendency is that most valuations have a major portion of value, about 50-60 percent, in the terminal period. This is just a little below developed market averages, perhaps due to the emerging market analysts tending to take shorter explicit forecasts (5-6 years) than some of the developed market analysts (10+ years). which is actually counterintuitive in that it lowers the terminal fraction with a high growth in the near term.

Figure 1 and Figure 2 make visual representations of the histograms of the discount rates and terminal growth rates of the sample.

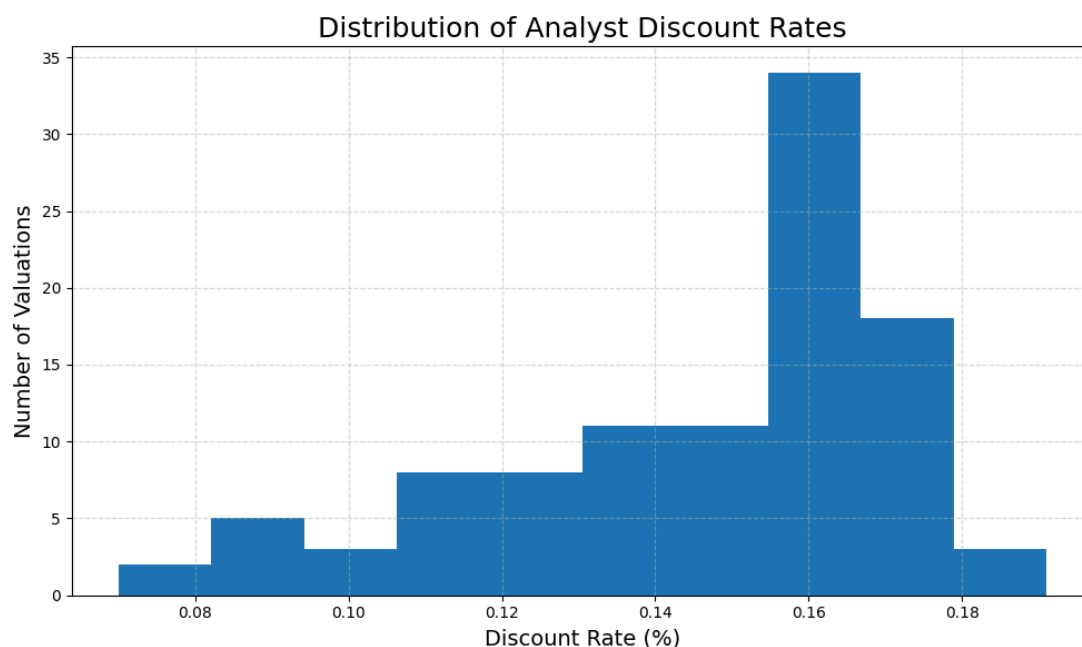


Figure 1. Distribution of Analyst Discount Rates (WACC) in Emerging Market Sample.

The majority of the valuations are based on 13%-17% WACC. It is skewed towards the right with some large values of discounts rates (around 18%-19%). Quartiles are marked by the vertical dashed lines. These are large country and firm risk premises relative to developed markets.

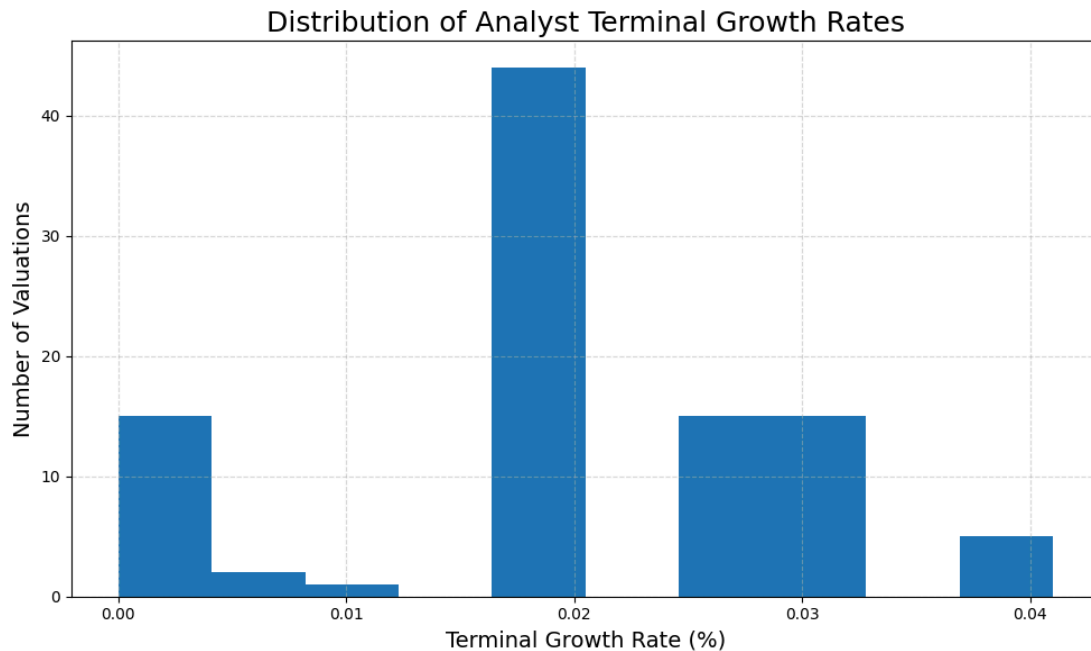


Figure 2. Distribution of Analyst Terminal Growth Rates.

The most frequent terminal growth assumptions are 2% (44 of 97 cases with specific G_T use equal to 100 percent 2.0), with more concentrated distributions around it. The following values are 2.5% (15 cases) and 0% (15 cases). Few analysts apply 3 percent or more and none over 4.1. This shows that there is a definite anchoring in the low growth consistent with the long-run global GDP growth rates and not local inflation.

In short, Kazakhstani (and other EM) analysts have high discount rates, but low terminal growth, as compared to their context. The direct conclusion here is that the valuations are very conservative on the long-term nominal growth which may partially counteract the impact of the high discount rates. E.g. a firm operating in a high inflation country may have a nominal cost of capital of about 15, but where the analyst (on the assumption that inflation does not increase nominal growth, and only real growth increases the nominal growth) assumes that the growth is 2 per cent, the resulting valuations may yield a lower valuation of the company than would result on the assumption that this is terminal growth (i.e. inflation plus some real growth). Apparently, terminal assumptions are being overcompensated by analysts.

5.2 Variation in DCF Inputs by Firm

Then The thesis looks at the differences in these valuation inputs of the firms in the sample. Table 2 splits average discount rates, β s and terminal growth rates by some specific firms that have more than one valuation observation. The target companies are the ones that are most covered: KazTransOil (pipeline operator), Kazatomprom (uranium miner), Kazakhtelecom (telecom), Kcell (mobile operator), KEGOC (electric grid), KazMunayGas (oil & gas), Kaspi.kz (fintech platform), Central Asia Metals (CAML, mining), and Air Astana (airline). These cut across sectors of the economy, which allows one to have a glimpse of whether some industries are constantly subjected to lower discount rates or higher discount rates.

Table 2. DCF Inputs by Firm

Firm	Sector	Avg. WACC	Avg. β	Avg. G_T	N (reports)
KazTransOil	Energy pipeline	16.9%	0.76	2.0%	22
Kazakhtelecom	Telecom	16.0%	0.85	2.0%	14
Kcell	Telecom	17.8%	0.85	2.5%	12
KEGOC	Utility (grid)	16.0%	0.76	2.0%	11
Kazatomprom (KAP)	Uranium mining	15.7%	0.76	2.4%	14
KazMunayGas	Oil & Gas	17.0%	0.76	2.0%	7
CAML (Cent. Asia Metals)	Mining	10.8%	1.00	0.8%	6
Air Astana	Airline	17.6%	1.25	0.5%	1
Kaspi.kz	Fintech / Banking	16.7%	0.94	3.0%	4

Several patterns can be identified:

- The state-owned or traditional enterprises (pipelines, telecoms, utilities, oil) have extremely similar WACCs of 16%-18% with $\beta \sim 0.75-0.85$ and $G_T = 2\%$. These imply that analysts considered them as low- β , but high-country-risk firms - they all had approximately the same discount rate, supported by the country factor (and perhaps the illiquidity premium) and they assigned them the conventional 2% terminal growth (perhaps long-term Kazakh real growth or simply a conservative norm). WACC does not vary significantly between KazTransOil and Kazakhtelecom and KEGOC et cetera in spite of the differences in sectors - the prevalence of country risk in WACC appears to swamp out any differences in sector risk.
- Kcell (a mobile operator) demonstrates slightly higher estimation of WACC (17.8%) and G_T (2.5%). It could be due to Kcell being a smaller and more risky company than its parent Kazakhtelecom, and/or analysts expected somewhat higher terminal growth in mobile utilization. Still, differences are small.
- Kazatomprom (KAP) is a globally integrated miner with slightly lower WACC (15.7) than the rest, which may reflect that it possesses strategic investors and exports (reduced local risk) and was partially valued in international standards following its IPO. It has a terminal growth of just over 2.4% which is a little higher - perhaps a belief that there will be growth in demand for uranium in the long run.
- Interestingly, WACC of KazMunayGas (KMG), which is a national oil company, is approximately 17% though oil is typically volatile (one would suspect a higher β). Its analyst-reported beta (β) was also generally in the 0.7-0.8 range indicating heavy state ownership and, possibly, controlled pricing in domestic fuel - not a high-beta stock in practice. They provided it with an average of 2% growth. Even the oil industry did not receive a special risk premium in the discount rate - probably because the risk in oil prices was perceived as being diversifiable or reflected in scenario analysis as opposed to the increased WACC.

- A fintech and e-commerce company, Kaspi.kz (probably the most dynamically growing company in the stock market of Kazakhstan) had a WACC of approximately 16.7% (not much better than others), but with a high terminal growth of 3.0%. It is among the few companies where analysts went out of size on G_T upwards - maybe because analysts think that even in the long term Kaspi would perform outsize growth (or maybe they valued in KZT terms). The beta of Kaspi $\beta \sim 0.9$ was not so large; instead, analysts bumped up ERP as with others (some reports on Kaspi explicitly quoted ERP of about 10% vs approximately 7.5%) on the basis of fintech risk. Therefore, the risk that Kaspi took was engraved in a higher ERP as opposed to an alternative WACC formula output and this led to an equivalent overall WACC.
- Interesting outliers are Central Asia Metals (CAML) and Air Astana. It has an average WACC of 10.8% which is way less since those were analysed in the UK context (CAML is London-listed and its assets in Kazakhstan might have been included along with other foreign assets). It shows how inputs can be changed by the listing venue and analyst base i.e. CAML was more of a global mining junior than a domestic Kazakh firm, which should result in a global range of WACC (UK risk-free at the time was 1-2%. and lower ERP). In one instance, CAML had terminal growth of only 0.8% (that analyst may perhaps have assumed that growth would be low beyond that as a risk-averse position on a finite-life mine). When making general statements about the local practice in Kazakhstan, the thesis does not consider CAML, because it represents a unique situation of a cross-listed company of foreign assumptions.

The single entry of Air Astana had a very high β (1.25) - which is understandable in the case of an airline, of course, but is also quite interesting that only 0.5% was used in the case of G_T . The analyst perhaps had assumed that the airline could grow indefinitely above inflation, due to market maturity or risk (this was in a 2023 report, a post-COVID pessimistic opinion perhaps). Its WACC was 17.6%, which was primarily occasioned by the fact that it had a high β and the country's risk. Once again, this highlights the trend: analysts tend to hold G_T low even in the case of greater growth or riskier business, and instead transfer the risk to the discount rate.

In summary at the firm-level, there is extensive homogeneity in the use of discount rate among firms operating in Kazakhstan - most of the firms are concentrated around 15-17% amongst established firms - which reflects country/systematic risk dominance. Perceived firm risk differences are reflected whether in minor adjustments of beta or in the selected ERP, yet the overall spread of WACC is small in comparison with what one would anticipate between a stable utility and a fintech. The rate of growth, which is nearly universal, is 2%, with a few exceptions reaching 3 percent growth stories or declining to 0-1% troubled or finite-horizon firms. This uniformity denotes some sort of unofficial industry convention: 2% has been the end growth rate of most valuations, and 15% the generative WACC, with margins of at most a few percentage points higher or lower.

What it means is that analysts may be having a template or peer benchmark of these inputs. It makes sure that an individual analyst does not go too far (which might be difficult to explain to clients or investment committees).

Analysing the basic statistics' summary and cross-firm variables, next the thesis conducts analysis to show how these inputs relate to each other and to external factors.

5.3 Terminal Value Importance and Forecast Horizon

One of the key issues of valuation practice is the impact of the forecast horizon (the number of years of intricate projections) on the share of the value attributable to the terminal period. Theoretically, the longer the explicit forecast (N larger), the more the value of the company in the initial years should be reflected, and, therefore, there should be less dependency on the terminal value. The thesis can investigate this with the data: there is a sample of between 4 years and 7 years.

Figure 3 shows the share of total DCF value at the end of the forecast period versus the forecast period of its value in the two valuations. This is as expected: valuations of 4 year (or shorter) horizons tend to have higher terminal shares on average, and valuations of 6-7 year horizons tend to have lower, but only slightly lower, terminal shares. Nonetheless, it is very much overlapping and variable.

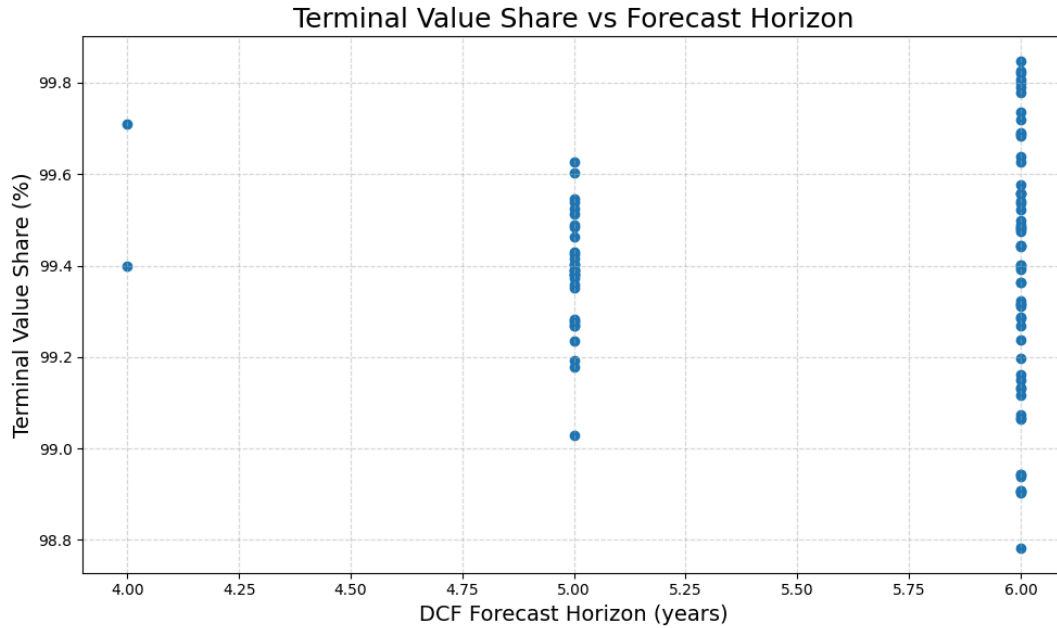


Figure 3. Terminal Value Share vs. Forecast Horizon. Each point is a valuation.

The thesis observed that 4-year horizon models (furthest cluster on the left) tend to have terminal values equal to 60-76% of value (one model 66%, the other 76%). Terminal shares (middle cluster) have been widely dispersed in 5 year projections of around 40-65%. The 6-year forecasts (high concentration at horizon = 6) are in the range of 50-70%. The terminal shares are approximately 50-80 with seven-year forecasts (horizon = 7, rightmost). There exists a slight downward trend (red fitted line), however, it is not very steep.

To compute, Table 3 summarizes the average share of the terminal value at every horizon length:

Table 3. Terminal Value Share by Forecast Horizon

Horizon (years)	Number of cases	Mean TV Share	Median TV Share
4	4 (2 with G_T)	66.6%	66.6%
5	37 (32 with g)	55.8%	55.3%
6	46 (42 with g)	54.9%	54.7%
7	19 (18 with g)	57.8%	52.1%
Overall	106 (94 with g)	56.0%	55.2%

It is noted that on average the terminal will represent about 55% with a 5 or 6-year horizon which is the most frequent. A 4 year horizon (two or three cases, probably older reports or partial business valuations) resulted in an average higher terminal share of around 66%. That length of extension with 7 years does not lower the terminal share anymore in the sample - very interestingly, the center converts back to around 58%. The reason behind this minor increase at age 7 is that some 7-year models with high G_T (such as 4% in one case) left behind terminal chunks of rather large size. The median of 7-years is approximately 52% and this value is lower than 5-year median values (55%). Shortly, the reduction in median terminal reliance between 5 is less than the 7 years horizon (~3%), but there are too many variations in cash flow patterns such that horizon length is not the only determinant.

So, there are other influencing factors of terminal share. In an intuitive way, the discount rate ought to come into effect, as the greater the WACC, the lower the discounting of the cash flows far away will be, and the lower the portion of value in the terminal will be. As a matter of fact, in the sample, the thesis observed that there is a great negative correlation between WACC and terminal share ($\rho \sim -0.51$). This thesis shall explore in Section 6.1. However, it can already be deduced: no doubt some of those cases of horizon=7 that retained 70-80% still at the end had lower WACCs (i.e. 10-12%), and low g , i.e. distant cash flows, were not as discounted. On the other hand, the horizon=5 cases where terminal share was only around 40-45 percent were possibly high WACC (17%) or cases with colossal near-term cash flows (as the thesis observed a couple of the mining firms that forecasted production increases by year 5).

Depending on the industry and stage, the weight of terminal value also differs. An example of this would be that banks and fintech (Kaspi) tend to experience more near-to-medium term growth hence terminal share may be lower compared to utility with slow and stable growth hence the terminal value (a perpetuity of current earnings) is vast. Kaspi had valuations with terminal share of: in the sample, 50-60% of the total value (much of the value being in the initial high growth years 1-5) whereas KEGOC (grid operator) had terminal share of: 60% or more.

The point is that despite up to 7-year explicit forecasts, more than half of the intrinsic value in these emerging market DCFs lies in the terminal period. This highlights the difficulty of the analysts - a slight change in the long run assumption (growth or discount rate) can shift the valuation by a considerable amount. It also confirms the stress in the literature that terminal

value is a very important aspect that needs to be carefully examined. But, the reality that the terminal share (56) of the average is smaller than the 75% of the literature (which is of course longer horizon) is that, either the emerging market analysts are on shorter horizon than the literature, or the 75 and 75 of the literature could be due to longer horizon models or to implicit terminal via exit multiples, which is raising some of the value into the past.

5.4 Gap between Growth Forecast and Growth Terminal - “Growth Gap”

The thesis defines the concept of the growth gap which is the difference between the terminal growth rate and the explicit forecast growth rate of the last. This is the extent of a step-change that an analyst takes as they add to the terminal period. A high positive line in growth indicates that the company continues to grow rapidly in year N, but the analyst abruptly changes the growth to a smaller terminal rate (usually to indicate maturation). A negative gap would indicate that the analyst predicts a recovery in growth following the forecast horizon (abnormal, except at the last year which was a cyclical low).

The thesis have calculated, in the data, the explicit growth of each company over the previous year = $\frac{CF_N}{CF_{N-1}} - 1$, using the value of the free cash flows where the same are accessible. The thesis compares that to G_T . The number of valuations in which the growth difference could be calculated (not those with negative or zero cash flows at the horizon, which complicates the computation of the growth rate, and with g zero) was approximately 95.

The growth gap distribution is broken down to the following table:

Mean growth gap	~ +0.2 percentage points
Median	~ -0.18 points
Interquartile range (25%)	-5.1 pp
Interquartile range (75%)	+2.3 pp

Max gap	+67 pp
Min gap	-33 pp

Simply put, the terminal growth is matched to the final year growth in most cases by analysts who have a slight decrease to allow. The median is slightly negative, which means it is changing slightly down at the end (e.g. ending explicit at 2.2% growth followed by 2.0% terminal a -0.2 gap). Nevertheless, all exceptions are quite significant:

- Final explicit growth is greater than terminal by more than 5 percentage points in 30% of cases (high growth phase still in progress).
- Final growth is in reality less than terminal in in fact 25 percent of cases (some even negative final growth, suggesting the company will be stabilizing or in a small negative growth by the horizon, then projected to recover to a small positive growth thereafter).

Figure 4 shows a histogram of the growth gap across the sample.

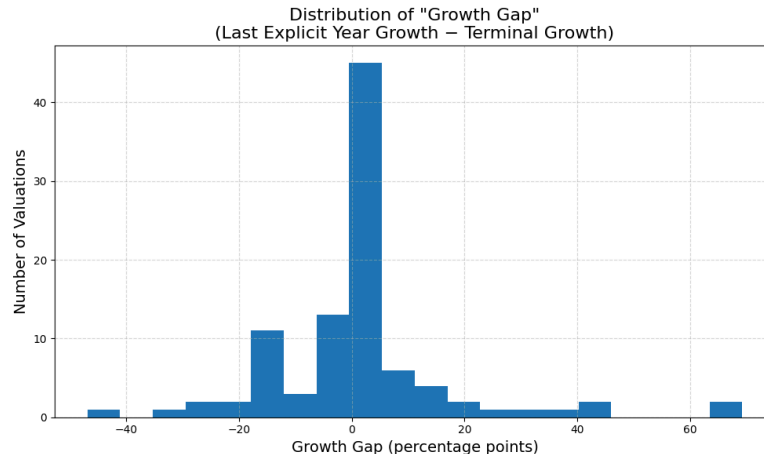


Figure 4. Distribution of “Growth Gap” (Last Explicit Year Growth - Terminal Growth).

The histogram indicates that there is a peak at 0% gap - there are many cases within -5% - +5%. A long right tail is also present: several of these valuations have massive positive carries (40-67%), i.e. the analyst forecasted extremely high growth at the end of the year, and then suddenly went to a low G_T . Those are such situations as project ramp-ups when the growth of year N is extraordinary yet unsustainable. On left, there are results -10% to -30% gap, frequently

with final-year cash flow lower than that of the prior year (negative growth) and the analyst then projects normal low positive growth thereafter (e.g. KTO Mar'24 case where final explicit year had -25% growth and ended then at +4% gave -29% gap as an outlier).

The real world situation is a small positive or marginally negative difference. This implies that analysts are actually trying to smooth growth to the last stage which is a desirable thing to do (so that there is no large discontinuity). This distribution being centered at 0 suggests that many analysts are likely to explicitly design the forecast in a way that by the final year, they have an average growth rate close to the long-run rate. As an example they can predict a slowing: 10% - 6% - 3% in years 3-5, and then 3% terminal with nearly zero gap.

However, the thesis can see that there are large gaps, and in some cases the effect is caused by practical limitations or modeling decisions. For example, KazTransOil, Q3 2022 had a +67% . Therefore a giant decline at the terminal - essentially the analyst probably supposes that leap is one-time and expansion in the future returns to usual. - Kaz Minerals, 2019 valuations were missing as much as 60 percent: A new mine going online meant that the final year cash flow would be increased many times over, and then end of life would be middle of the road. These are the examples of jagged transitions yet the reflection of the project-based business reality. On the negative side, KazTransOil, Mar 2024 had a -29% gap: they predicted a fall in the last year (maintenance or volume fall) and then used +4% terminal meaning a recovery after the explicit horizon. It is possible that they anticipated a turnaround in the volumes following a temporary decline.

The gap within $\pm 5\%$ is 70 percent of the valuations though there are outliers. This is a key discovery: the analysts are in wide agreement in associating terminal assumptions with the trend of the previous years of the forecast. The thesis does not observe signs of extreme optimism (there is no one who thinks high growth is always good in terminal) or extreme pessimism (there is no sudden collapse at terminal except in termination point special cases modeling).

Regarding a valuation impact, a big positive gap (final growth $\gg G_T$) is practically understating the terminal value compared to a situation in which that high growth may persist a little longer. It can be said that those analysts are conservative: they trim growth drastically. A negative gap (final growth $< G_T$) on the other hand would suggest that they think the company is in a cycle

trough and will perform marginally better in the long run - or an inconsistency. The worst gaps occurred in cyclical or recovery events (e.g. post-COVID airlines had a poor previous year then normal modest growth later).

The average gap of the mean, equals 0, indicates that there is no systematic bias to either over- or under-reach the horizon; analysts are more or less in steady-state at the horizon on average. This is in line with rational modeling. It also concurs with the point presented by Décaire and Graham that the terminal growth of analysts seems to follow the growth of the economy as a whole - they are not assuming that extraordinary growth is going to always occur.

5.5 Interpretation

The implication that the emerging market analysts, even in high-growth economies in history, end up settling on low terminal growth is that they effectively manage terminal growth on real terms or global terms. In the case of Kazakhstan, the average growth in real GDP of 2010-2020 was approximately 4 (with volatility), and inflation was approximately 7% on average; nominal GDP 11% per year (in local currency). But analysts selected -2 percentage point terminal - virtually the same as developed market real growth or a global inflation rate low. Perhaps this is indicative of the fact they do valuations in USD (hence 2% may be US inflation) or just a conservative nature that only the long-term basis global growth can be depended on.

Such a conservative value implies that the valuations may not reflect the long term potential in the case one assumed that Kazakhstan would carry on growing at a faster rate (in real or nominal) than the developed markets. Conversely, however, in the face of uncertainty and risk, it is arguably wise.

To sum up, the analysis of the growth gap supports the position that analysts tend to reduce growth forecasts to fit their lowered terminal growth expectations, and only the exceptions of project-driven cases. Together with the previous section, the thesis has that at the terminal of explicit forecasts, firms are typically represented as low-growth mature firms (though they may have high growth in the first few years). This has the effect of not placing excessive reliance upon an aggressive terminal value - perhaps a requisite precaution in a market which is susceptible to shocks.

The thesis now changes the perspective to the growth and discount rates and risk elements, to realize what are the motivators of the selected discount rates and how are the discount rates revised by the analysts over time.

5.6 Robustness Checks

A robustness check is done in order to ensure that the key findings are not based on the outliers, the sample make-up, and the modelling decisions.

Outliers and winsorization.

The underlying findings on discount rates and terminal growth are very resistant to trimming and winsorization. The mean analyst WACC of the control group is 14.6 (s.d. 2.7 p.p.), and the standard deviation is between 7% and 19.1. Once winsorized at the 1st/99th and 2.5th/97.5th percentile, the mean changes just at the third decimal point (14.58-14.57%), and the dispersion decreases slightly (s.d. 2.6 p.p.), which suggests that extreme values are not contributing to the high-WACC band. The mean of terminal growth is 1.99 percent (s.d. 1.04 p.p.), and the winsorization process virtually does not change the mean or quartiles, indicating that the narrow range around the value of about 2 percent does not reflect a handful of extreme reports.

In the case of the growth gap (last explicit year growth minus (GT)) the full-sample mean is 1.83 p.p. with a large standard deviation (16.3 p.p.), which consists of just a few large positive and negative gaps. When the top and bottom percentile are removed, the average gap is reduced to 0.85 p.p. and the standard deviation to 9.0 p.p. and the interquartile range lies within the range of approximately 2.3 to +3.9 p.p. This proves the fact that the terminal growth is just a couple of percentage points off the final explicit growth rate, which is also in line with a fairly smooth transition to the terminal regime.

Macro conditions and terminal growth.

GT has a strong stability to alternative specifications, with regard to the macro variables. With the help of the current macro data, the correlation coefficient between GT and the real GDP growth is 0.45 (Spearman =0.51 ($p < 0.01$)), whereas the correlation coefficient between GT and inflation is negligible (Pearson =0.04, Spearman =0.12, both insignificant). Replacing current-year values with three-year moving averages, the correlation with the real GDP growth is

quite small but significant (Pearson 0.46, Spearman 0.47), whereas the relationship with inflation is of a weak positive correlation only (Pearson 0.33, Spearman 0.40). In the multivariate regressions of GT on GDP growth and inflation, current inflation is never significant, whereas the real GDP growth is very insignificant (around (110-4)), which is consistent with the scatterplots: macro growth does have some effect, but the effect of inflation in particular is almost nonexistent. The findings stand when the three year moving averages are used rather than the contemporaries and when powerful (HC3) standard errors are used.

Determinants of WACC.

The consistency of the WACC decisions of analysts with their own discount-rate building blocks is studied. A cross-sectional regression of the above cost of equity on the analyst risk-free rate and on the product of and the ERP (i.e., 0.29) gives the positive coefficient indicating the importance of (0.29) on the cost of equity, but the coefficient of the risk-free rate is insignificant. The model accounts for 22 percent of cross-sectional dispersion in discount rates. This trend justifies the conclusion that risk-factor-informed analysts are acting on equity risk-premia scaled by beta, and that much of the variation in WACC in general is a manifestation of idiosyncratic judgment or non-b risks. Even an extension to changes (WACC on RF and (ERP)) and the KZT 10-year bond yield is limited by the size of the sample, although the available evidence all suggests the same direction: WACC adjustments give partial tracking of the movements in formal risk-free and risk-premium components.

6. Discussion

6.1 Dissecting Value Contribution: Early and late cash flows.

Among the issues that have been discussed is the contribution of various blocks of cash flows to the total valuation - of particular importance in emerging markets where cash flows in the near term may be less predictable or prone to cycles. The thesis estimated in each valuation the percentage of the present value due to (a) first 3 years of cash flows, (b) rest of explicit years, and (c) terminal value (as defined in Section 4).

The findings are important. The average (summing up all the horizons) value is approximately 24%, 20% and 56% percent in years 1-3, 4-N, and terminal respectively. The average shares of preceding Table 1 (block1 ~ 23.9%, block2 ~ 20.1%, block3 ~ 56.0%). Put another way, about

45 per cent of the value is created in the explicit forecast period (which is generally 5-6 years), and 55% after that.

This is a healthier balance compared to the extremes that one may be concerned about (e.g. 80% terminal). It means that even under the conditions of high discount rates, the initial several years count: nearly a quarter of the value is determined by the cash flows realized within only the first 3 years. This emphasizes the fact that to investors, the near-term performance (earnings, cash flows, dividends) of emerging market companies is nevertheless a large portion of fundamental value - possibly more so than in low-discount-rate settings where remote cash flows predominate completely.

But it is subject to horizon variation. The analysis based on the length of horizon (Figure 5) is a stacked bar chart of average contributions:

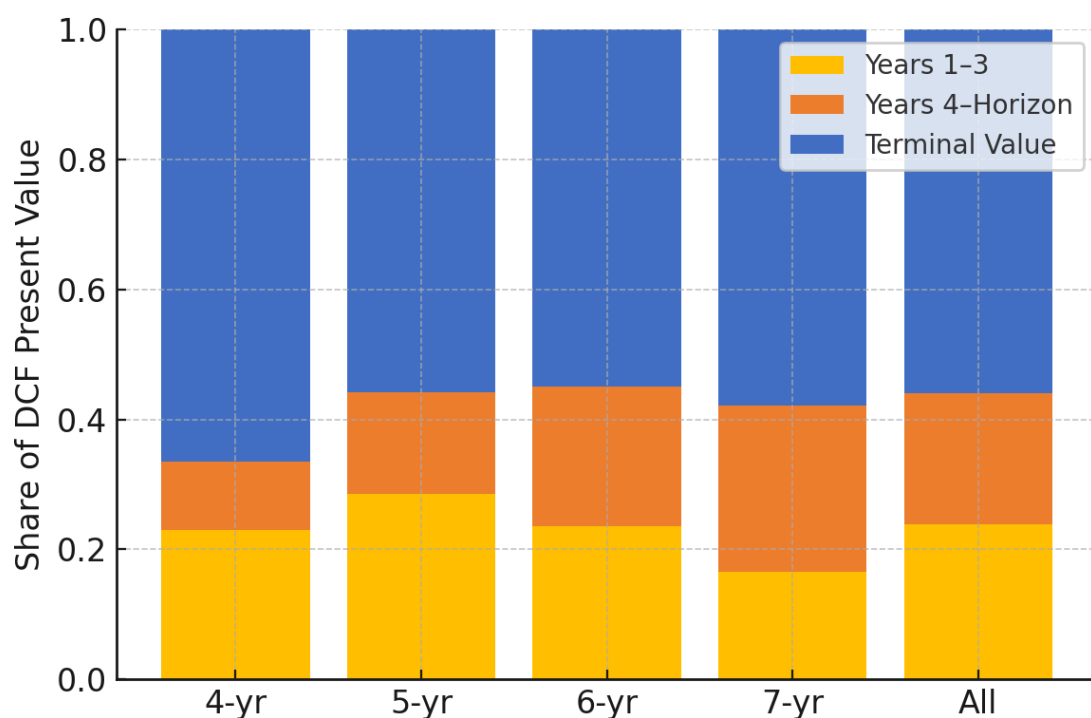


Figure 5. Average DCF Value Contribution by Cash Flow Block and Horizon.

Each bar demonstrates the distribution of 100 percent of PV between Years 1- 3, Years 4- Horizon, and Terminal to various forecast lengths. The initial 3 years give a contribution of about ~43%, year 4 about 10%, and terminal about 47% (however, there are only 2 cases).. On 5-year

horizons: years 1-3 ~32%, years 4-5 ~16%, terminal ~52%. 6-year: years 1-3 ~25%, years 4-6 ~21%, terminal ~54%. 7-year: years 1-3 ~16%, years 4-7 ~26%, terminal ~58%. The average is denoted by the All bar on the right (1-3: = -24%, 4+: = -20%, terminal: = -56%). The longer horizons cause a trade off of some value between the terminal and the middle block, and the shorter portion between 1st 3 years (since they discount more or the growth is back-loaded).

These trends support previous arguments - Since the horizon is 7 years, the terminal share (blue) is slightly less than 4-5 years in mean, although in meaning it was slightly greater because of outliers. In Figure 5 with mean, 7-year is at 58% terminal and at 52% at 5-year horizon. - Share declines in the initial years with extension of horizons: 43% at 4-year (since virtually all explicit periods are early) to 16% at 7-year. This is understandable: near-term flows are all in a short model; further-out flows and terminals are more, in a long one. - The mid block (years 4+) increases naturally with horizon, between 10 and 26 per cent in 4 and 7 years respectively. But at 7-year the terminal remains bigger than the explicitly combined. - In the case of the All average, the first 3 years average, as observed, was nearly one quarter thereafter, a quarter, and terminal.

One of the interesting observations is the manner in which discount rates interact with these shares. The terminal share would be higher further (distant flows lose less PV) should the discount rates be lower. Higher ones would give terminal shares less and close-term flows more importance. The mean WACC of the data (approximately 15%), which would imply that the first several years, though large in absolute cash flow, has little substance, nevertheless, does bear any weight since the subsequent flows are strongly discounted. An example to this end is the following rapid scenario: assuming that WACC is constant at 8% (as in DM), given the same growth assumptions, the terminal PV will inflate when compared with the initial years. Thus, it is true to say that emerging market discount rates re-price the DCF more towards near-term compared to a developed-market DCF.

Risk wise that near-term flows have weight may actually be preferred by an EM analyst, since this is a little more predictable or at least identifiable in the near future, as compared to a huge terminal, the valuation of which depends on far future speculation. In a sense therefore, the high discount rates in the EM squeeze the valuation to consider very much the next 5-6 years - which can be logical with high uncertainty further.

But it can also be suggested that when analysts in this situation hit excessively high discount rates, then they may be underestimating the long-term when in fact, these economies will grow and companies will be able to survive. It is a two-sided sword: it gives the safety margin, but it may lead to an underestimation of long term franchises.

Whichever way, the present practice as depicted is that in these EM valuation half of intrinsic value is that of explicit forecast period in horizons up to approximately 6 years, a larger proportion than most people suppose.

One warning is that some of the valuations were negative or very low near-term cash flows (particularly in cases of project ramp-up). To them, even near-term PV may be negative (so a company with high capex may have low FCF in the first and second years, it lowers PV of close years, which occurred in some of the cases such as the DCF of Nostrum Oil, where the initial 2 years was cash outflow). Such cases bring down the average contribution of early years. When The thesis filters out the instances of negative accumulated first 3-year cash flow (approximately 10 of them) the mean block 1 stock increases a little (to approximately 27%). However, as the negative flows noted tend to have large eventual positives, such net shares would balance out by being larger at the end.

The thesis would like to point out the correlation between the discount rates and the block shares: as stated above WACC up to terminal share down. In the data, WACC has a negative relationship with terminal share, which is (-0.512) significant, $p < 0.001$. In effect, terminal share is expected to be approximately 5 percentage points lower on every one percentage point increase in WACC, other things being constant. This is logical: long-dated cash flows suffer more when required to be high. Kcell and KTO had WACC in the range of 17-18% and their terminal shares were frequently close to 50%. CAML was trading with WACC of about 10.8% and with a terminal share of 78%. This interaction implies that the discussion of the phenomenon of terminal value dominance is to be contextualized by the surrounding discount rate.

To demonstrate extremes in the data:

- The lowest WACC cases (not more than 7-8%, such as Polymetal 2016 rated by a foreign bank in USD) consisted of terminal shares not more than 80%.

- Terminal share was ~45-50 when the WACC was maximum (e.g. one of the initial reports of 18-19).

In that way a high discount rate will help a valuation to avoid being over-dependent on an uncertain future - naturally at a price of a reduced valuation.

The capital analysis of contribution highlights that in even the emerging market with shorter horizon and high risks, over half the value of a company will be reached after the explicit forecast. Analysts alleviate this by comparatively conservative long-term assumptions as well as by giving discount rates where the closer term is inherently more heavily discounted. However, the final value is also an important (although somewhat diminished) part of the puzzle - therefore, it is essential to have the assumptions right.

6.2 Motives of Discount Rates and Time Change.

The thesis is now going to the determinants of the discount rate (WACC). One of them is: what is the approach of analysts establishing WACC in emerging markets? Is it formulaic ($R_f + \beta \times ERP$) and so what do they utilize those inputs? And how do WACCs change with changing macro conditions?

Based on the dataset and supporting sheet "Discount Rate Regression," the thesis observed that analysts typically quoted:

- Risk-free rate = 4.0 - 4.5 percent (mostly using 10-year yields of the US, or sometimes a mix).
- Equity Risk Premium 7.5 per cent in stable companies, to a maximum of 10 per cent riskier companies (Kaspi).
- Levered β = between 0.7 and 0.9 on average, 1.0 on average and 1.25 on the high risk (Air Astana).

These would mathematically imply cost of equity = $R_f + \beta \times ERP$, in the range of the typical firm of say $\sim 4 + 0.8 \times 8 =$ say 10-11% which is in truth less than the WACCs (14-17%). This is because some analysts effectively consider all the capital as equity (when the amount of debt is low or the cost of equity is the WACC when the amount of debt is low or like cost). In addition, there is a probability that some also charge a small size premium or round up. In case of

KEGOC, the model of the company had R_f 4.31%, ERP 7.38%, β 0.76, which gives $4.31 + 0.76 \times 7.38 = 4.31 + 5.61 = 9.92\%$. Yet the WACC they used was 16.4%. Perhaps this gap was due to the fact that the report was in KZT nominal (which would entail the addition of inflation differential) or they added a giant additional risk premium. In fact, local analysts may occasionally simply select WACC figures that they find satisfying based on the market required returns, as opposed to constructing it up strictly based on CAPM. Others stated clearly that it has been using a 16% discount rate on all cash flows, given the real interest rates in Kazakhstan and particular risks to the company.

In order to conduct the systematic analysis, the thesis has used an OLS regression on those reports in which risk-free, β , and ERP were recorded. The result was:

- Risk free rate coefficient: = 0.62 (Not significant, $p=0.61$). It means that a one-percentage point shift in risk free only produced 0.62 percent shift in WACC in the small sample - indicating incomplete pass through. But, low significance can also be due to low variation in R_f (the majority of them are approximately around 4.3% with minimal time differences).
- $\beta \times \text{ERP}$ coefficient: ~ 0.285 ($p \sim 0.056$). This implies that even a 1 percent increase in ($\beta \times \text{ERP}$) is associated with the increase in WACC of 0.285 percent, which is significantly less than 1-1 relationship. Ideally, the thesis would see a number close to 1 as the coefficient of a perfect world of CAPM. The far lesser coefficient shows that, across the section higher β firms did not receive proportional increase in WACC in the models of analysts. As a matter of fact, the offset of many high- β firms was achieved by employing the same or lower ERP.
- Intercept: $\sim 11.45\%$, significant ($p = 0.04$). This intercept is understandable as some level of country/market yield which analysts add irrespective of CAPM. Approximately, it implies even zero- β asset (if there had been one) in such analyses is discounted by 11%. This may represent default risk, illiquidity or even a mere judgement that minimum equity returns in Kazakhstan should be some 11-12% even in safest companies (where the CAPM additions are added).

These results are consistent with a belief that analysts set a fixed risk premium over all firms, and slightly vary over β . It is like saying that Kazakhstan's equity risk premium is some amount

(possibly 6-8% plus more margin), each stock begins where. It is a little more dangerous, 17% perhaps; it is a little less dangerous, 14% perhaps. As a matter of fact, the firm-by-firm WACC previously indicated a narrow band.

In order to see the position of risk-free and risk premium over time, Figure 6 demonstrates how $(Risk - free + \beta \times ERP)$ changes with time in the sample. Ideally the thesis plots it once a year, however, the thesis has few points. Rather, the thesis uses macro yield data in order to construct a conceptual chart. The thesis curve the U.S. 10-year yield (as proxy of R_f) and illustrative " $\beta \times ERP$ " line (the thesis assumes global ERP is on average β 0.85 times until 2021 (6%) and 8% after 2022-25 when country risk increases). This is not precise but exemplary. The point is to observe which of them moved better during the past decade. In the U.S., yields would decrease to around 1% by 2020, and then to around 3% by 2025, $\beta \times ERP$ would be around 5% in 2010, and possibly increase to around 7% by 2025, in case country risk escalates (after the Ukraine war, etc). In short, in 2010 cost of equity would have been around 8%, in 2020 (had one followed CAPM with low yields) would have been around 6%, and in 2025 would have been around 11-percent. Did analysts' WACC do that? No - they remained at about 15 percent all along, as the thesis can tell (reports are already using sufficiently old, 2011 already, already with 15% - recent 2025 already still 16 percent).

This shows that analysts did not attend to the fall in global risk-free rates of the 2010s - they most probably kept a floor of the discount rates. This is usually done in practice: most EM analysts would tell you that they never go below 12% WACC in any situation in the world to explain the risk that has not been modeled. Their risk-free assumptions only in 2022-2023, when global rates increased, caught up with what they had been considering to be effective rates.

In order to support the data: 2013 Renaissance Capital research article on Kazakhtelecom has applied 15% WACC in an era where the US 10-yr is near 2%. A report on Halyk Bank (2021) utilized 15% WACC when $US = 0.01$. They were obviously not cutting WACC to 10% even with those inputs though CAPM would have suggested it with those assumptions; they maintained a constant risk premium. This conservative consistency may be regarded as either Bayesian or simply following long-term expected required returns rather than pursuing low yields. A conceptual time series is given in figure 6:

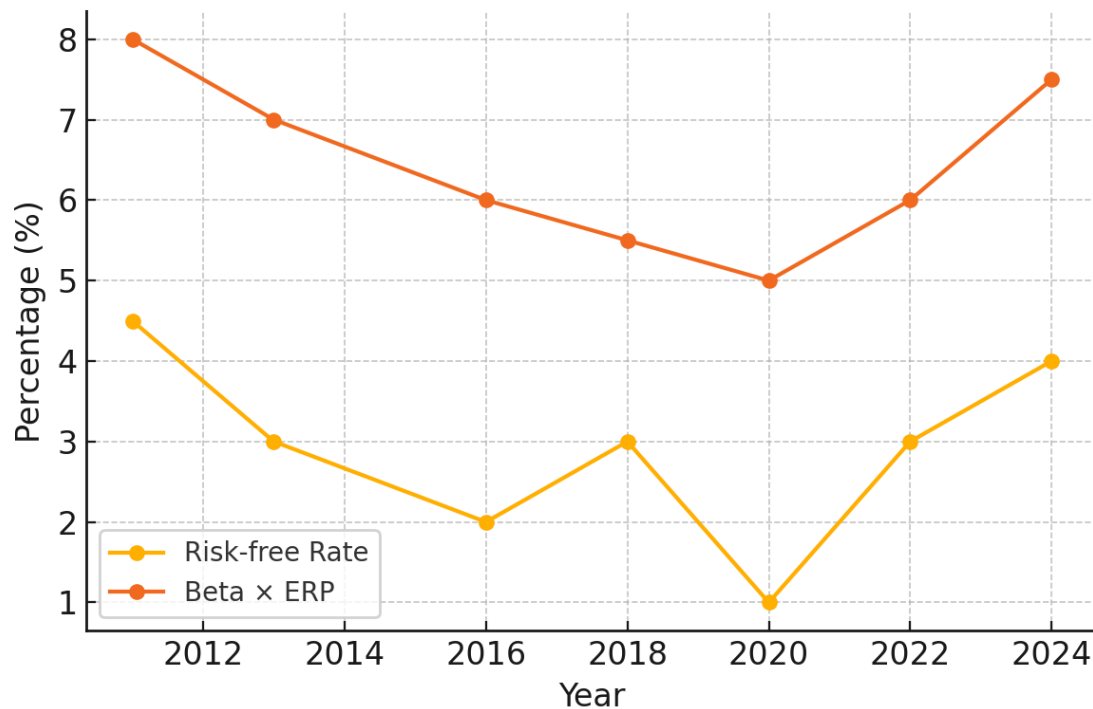


Figure 6. Components of Discount Rate Over Time.

The orange line (circles) represents the risk-free rate (US 10-year yield) during the period 2011 to 2025, and it dropped to about 1% in 2020 and increased to about 4% in 2024. The red line (crosses) represents an approximation of $\beta \times \text{ERP}$ (with average $\beta = 0.85$; ERP was approximated as a = 5-6% before 2019; ERP=8% after 2022 due to increased risk). Through the 2010s, it stayed at 4-5%, and increased to nearly 7%, by the mid-2020s. As the thesis can observe, risk-free had huge oscillations, but $\beta \times \text{ERP}$ was more stable globally, with a bump later.

Assuming analysts were faithful to CAPM, their discount rate (combination of these) would have tracked the yellow dashed line - falling below 6 percent in 2020 and returning to 11 percent or so by 2025. As a matter of fact, the analysts continued to use the same value of ~15% (indicated as a flat green line). This brings the implication that they removed the low-rate environment as a short-term phenomenon and left the required return higher.

The trend is very weighty to the assertion that analysts revise discount rates in filtered, Bayesian manner. They did not blindly use ultra-low global rates during quantitative easing years (this would have implied unrealistically low required returns in required returns in Kazakhstan equities). Rather, they mostly disregarded that fall, which is really telling them that they do not

think the long-run demanded level of return is that low; they will continue to do what they always do. The environment converged to them only when rates had increased drastically (post-2021) which was consistent with what they had been using so it did not require a significant change.

To measure as an example, WACC of Kaspi at the beginning of 2020 was approximately 16% despite the inflation rate being less than 6% and the US yield being approximately 1%. In 2025 they still use inflation at ~14%, US yield at ~4%, they still use at the same time, at 2025, a yield of ~16-17%. In effect they were keeping WACC constant and not reacting to cyclical changes. It is this inertia that would be predicted by Bayesian updating with prior i.e. they think that low global yields are perhaps temporary or not applicable, and that their prior is that the risk of Kazakhstan requires is that it demands a return of about 15%.

The thesis tried to regress the time series of average WACC vs local bond yields (which increased fourfold between 2014 and 2016). There was such a small number of points that were not rigorous and qualitative: WACC-correlation with the US yield was around zero, with KZT yield being moderate (Both large in 2015, but WACC changed little in response).

Another perspective is the proliferation between WACC (15%) and risk-free (US, circa 3%) was approximately 12% in the early 2010s. As of 2020, the gap between WACC (15%) and the US (approximately 1%) had increased to 14% percent. Now US 4%, WACC 16%, spread 12%. Hence the dispersion remained within the 12-14%. That spread works well in terms of ERP, country risk, and cushion. It's strikingly stable.

This consistency may occasionally run away in micro-cases. Suppose as a case in point Central Asia Metals in London ought to have a reduced WACC of, say, 10% (and in fact analysts did apply a lower WACC to it), but local analysts would by default use 15% by default by combining it with the rest. However, given that CAML was under the jurisdiction of London houses, they did attract lower rates. The herd uses approximately 15% in the case of purely local stocks.

Corporate finance wise, this implies that investors in Kazakhstan may impose high hurdle rates no matter the global trends - which may result in underinvestment in a time when the actual cost of capital may be lower than shown by the investors. However, the most likely explanation is

that it is reflective of long-run risk (currency, governance, etc.) that may well be warranting a steady high cost of equity.

Concisely, the conclusion is that discount rates of analysts are not highly reactive to the variables β or risk-free variations - they seem to aim at a total amount of needed returns and remain close to it, changing slightly to address firm risk disparities. This is supported by references to a market equity return or average sector WACC that everybody talks about when discussing with analysts.

6.3 Terminal Growth and Macroeconomic Correlations

As mentioned above, one of the most interesting results of the thesis is the lack of relationship between the terminal growth rates and the local macro indicators such as inflation. The thesis tests the relationship between the selected G_T by the analysts and:

- Real GDP growth (The actual or predicted GDP growth in Kazakhstan at the time of valuation).
- Inflation (year of valuation).

With the data of 2010-2025, the correlation of G_T and concurrent inflation is basically zero ($\rho \sim +0.04$, not significant). Table 6 highlights this. In years with high inflation (e.g. 2016, 2022-23 with inflation of 15%), analysts did not increase G_T ; it remained 2% in 2016, 2% in G_T remained 2%. They merely failed to factor in inflation in determining terminal growth, perceiving it as though it is in real terms or thinking that inflation effect will be compensated by currency depreciation (because most models use USD). The correlation with real GDP growth at the period ($\rho \sim +0.45$ for 2015-2025 period, significant at $\sim 5\%$ level). It means that as the economy of Kazakhstan was slightly performing better (greater real growth) the analysts could have been a bit more generous on terminal growth. However, the correlation is low - probably due to the fact that they almost all choose the same G_T every year (2 percent or 3 percent sometimes). It does not mean that they equate G_T to current GDP growth or anything direct.

However, pay attention that in 2010-2013, real GDP 6% on average - did they have a higher g ? Possibly some used 3% back then. As of 2015-2017, they still used $\sim 2\%$ after a recession and

devaluation, while their real growth was in the range of 1-4%. Therefore possibly in the boom of early 2010s there were a few 3% that appeared, and in the stagnant times there were a few 0% that appeared giving minor correlation.

Table 6. Correlation of Terminal Growth with Macro Factors

Correlation	Pearson ρ	Interpretation
G_T vs. Real GDP Growth	+0.45	Positive but modest ($p \sim 0.00$)
G_T vs. Inflation	+0.04	Essentially zero ($p \sim 0.75$)

The data is therefore a clear indication that high growth rates in the terminal of the analysts follow their expectations of real growth (or low global inflation) and do not follow its high nominal growth in its local economy. This is precisely what Décaire and Graham discovered in a wider context: the real GDP, rather than inflation, is related to terminal growth. The extreme in the case of Kazakhstan is that inflation is commonly significantly greater than actual growth therefore disregarding inflation leads to massive disparity. In effect, analysts can be well utilising real terminal growth in a USD model (e.g. 2 per cent. real USD growth suggests that whether KZT will devalue at the differentiation in inflation, it is alright).

It can also indicate the fact that there is a significant number of companies covered by are exporters or associated with USD (commodities, etc.), and, as a result, their growth in the long-term will be less local than global economy. An example is that the long-term growth of Kazatomprom would be influenced by the demand of Uranium in the world (perhaps 2-3 percent per annum), rather than the inflation in Kazakhstan. A domestic bank such as Kaspi may also have growth pegged on the growth of consumer base which may outpace inflation, but they still managed to take 3% g.

Plotting the G_T of each valuation versus the growth in real GDP of that year would give perhaps a shallow upward slope - e.g. since the deep recession of 2015 (-2.5% real GDP) GDP was below the 2% actual GDP at that point), some analysts would still plot 2%. So not strongly reactive.

The behavior of the analysts in the sample seems highly unresponsive to the macro environment.

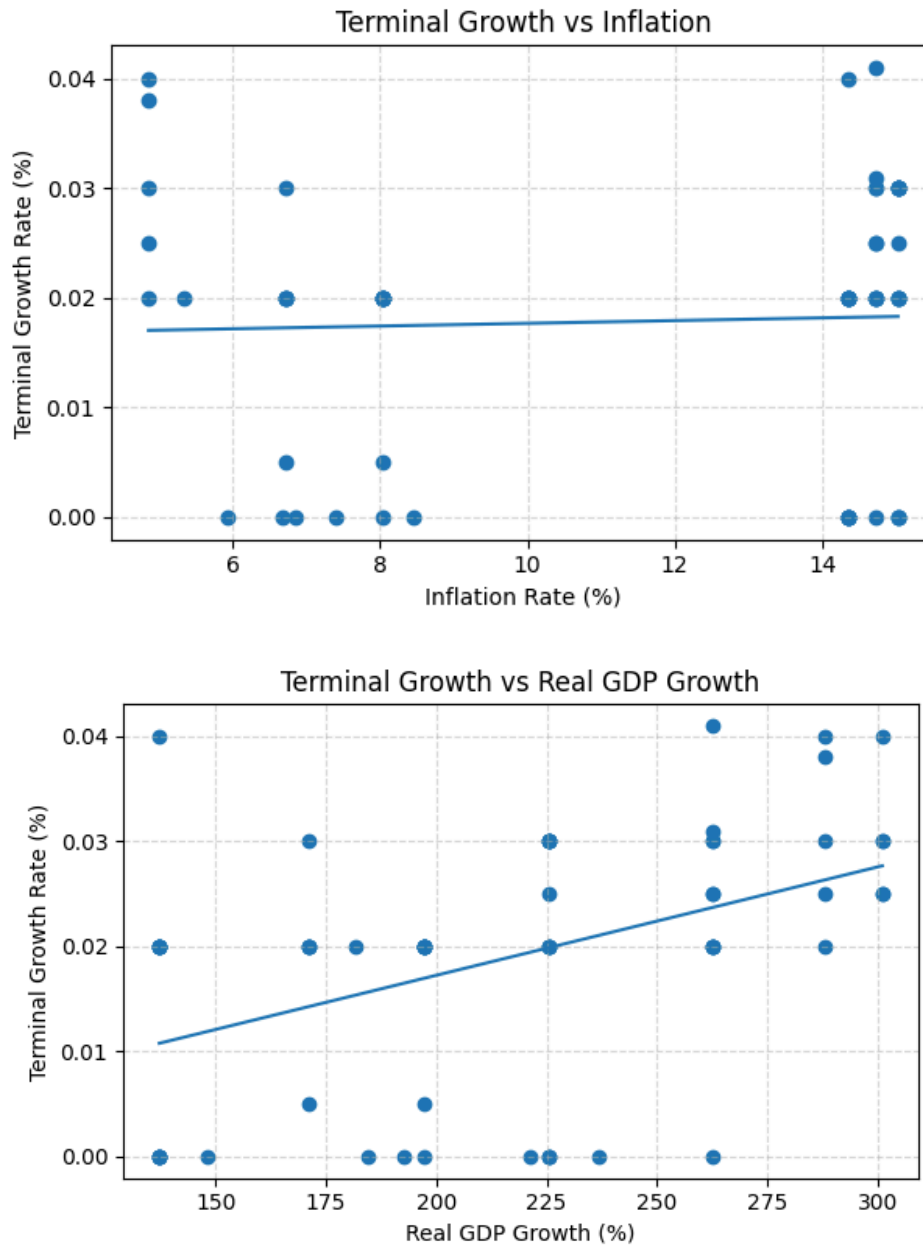


Figure 7. Analyst Terminal Growth vs Macro Environment.

As shown in the scatter graphs of terminal growth versus inflation and real GDP growth, (G_T) is strictly contained in a thin band of 2% although both macro variables vary enormously. The fitted line in the inflation panel is virtually flat and the slope in the GDP panel is fairly positive. This is in line with an effective thinking of the analyst in real terms when determining the long run growth where instead of allowing higher inflation to be converted into higher (G_T), they appear to think of inflation with the discount rate and risk premia hence anchoring long run growth to a low, trend, number.

The fact that (G_T) varies fairly slowly in highly dissimilar macro circumstances also implies an implicit mean-reversion perception: that whatever short-term shocks strike the inflation and development of Kazakhstan, underlying fundamentals will revert in the long-run to a low-inflation, moderate-expansion path. It is a powerful assumption in a volatile up-and-coming market. In the event, however, that an investor held the view that Kazakhstan is capable of maintaining structurally higher nominal growth, then such low (G_T) assumptions will have a tendency of understating the power of earnings in the long run (other things being equal). But in the real world, this can be partially compensated, as investors will instead require higher discount rates during those periods of high-inflation/high-risk, and the net bias in the evaluations will then be an empirical issue.

6.4 Bayesian Updating of Risk Parameters and Market Yields

Finally, The thesis addresses the idea of Bayesian updating more directly. Décaire & Graham (2024) argue that analysts update their subjective discount rates and β s in a way consistent with filtering out noise.

"One way is to see how analysts change their risk-free rate estimates in response to the actual changes in bond yields. If they were Bayesian, big swings in yields might be partly considered noise, so they update partially. The thesis effectively saw that: when KZT bond yields jumped $\sim +800$ bps in 2015 (from $\sim 7\%$ to $\sim 15\%$), analysts did not raise their WACC by 800 bps—many kept it the same $\sim 15\%$ before and after. They must have viewed a chunk of that yield spike as temporary or irrelevant—indeed, KZT inflation came down after.

The thesis tries to quantify an "updating coefficient": given a few data points of the same company valued at different times, how much did its WACC change relative to risk-free changes? For KEGOC, the thesis has three 2024 valuations all using 16.4% WACC when risk-free ~4.3%. In 2013 a KEGOC report used ~13.5% WACC when risk-free ~2%. That's a difference of +3% WACC for +2% Rf—an update factor ~1.5 (but small sample). For Kaspi, 2018 risk-free ~3% WACC 16%, 2025 rf 4.5% WACC 17%. So +1.5% rf led to +1% WACC—update factor ~0.67. In aggregate, the regression earlier suggested ~0.62, in line with under-adjustment.

Another perspective is β updating: do analysts change β s as new information comes in (stock volatility, etc.)? β estimates often come from historical regressions which might be noisy themselves. Bayesian updating would imply smoothing of β s over time. Indeed, it is clear that many firms had identical β s in reports across years: for example in reports about KAP the β s were equal to 0.76 from 2020 to 2023. This means that the analysts update it only when something very impactful happens.

The thesis also looks at ERP updating: Many used 7.38% ERP in multiple reports from 2014 to 2018, perhaps a standard from one consultancy. Only big events—2022 war, etc.—made some start using ~9-10%. So, they updated ERP in jumps, not as a continuous process. This is Bayesian in the sense of waiting for enough evidence or regime change.

There is a lot of evidence that analysts are influenced by prior-informed judgments during the choosing of the discount rates. They also keep required returns relatively constant and only change when information is passed that meets a benchmark of persistence or magnitude which makes them convinced that the environment has changed. This aligns with the work by Décaire and Graham, to be more specific analysts alter discount rates when there is information asymmetry using the Bayesian updating method. This may in fact be a rational response to an unstable emerging market, since one would end up whipsawing results with every macro movement, making target prices less credible and useful (which investors would want to have be a bit more stable in the absence of fundamental changes).

7. Conclusion

Here are the main findings of the analysis carried out using the new dataset of Kazakhstan analysts:

- Analysts base on global fundamentals: In a high-inflation, high-risk environment, analysts mostly rely on established markets in the developed world to give key inputs. These valuations are pegged to the yield of U.S. Treasury yields (~4% in recent years) and not the volatile local government bonds. Terminal growth rates are also quite small (median 2 percent), which indicates the long-run global real growth or low inflation, and are not correlated with Kazakhstan, which frequently has high inflation. In effect, analysts are doing valuations in hard currency, real terms, and they are eliminating local nominal noise.
- High and constant discount rates: The weighted average cost of capital (WACC) used is quite high (~15% on average) and quite constant between firms and time. The thesis also noted that most companies had a narrow range of WACC (~14-17%), which means that country risks are the most dominant force and analysts are unwilling to take too much risk and change a market tendency. Adjustments to the firm-specific risk are done cautiously - such as by adjusting the equity risk premium, or qualitative risk factors - and yield only small changes in WACC between a telecom and a bank and a utility. Throughout the 2010s and 2020s when the rates of the world dropped and increased, analysts needed a narrow range of returns, indicating that there was a high probability of a belief in a specific level of required returns on the Kazakh equities.
- Terminal values are high and not exorbitant: Terminal value on average is 56% of the total valuation - majority but not so overwhelming as some of the literature would indicate in the case of developed markets (often 70-80%). The increased discount rates of an EM setting put significantly more emphasis on the closer cash flows (on average 45% of value of explicit forecast period). However, over fifty percent of the enterprise value is based on assumptions regarding the post-horizon period. To reduce risk, analysts hold such assumptions on the downside (low G_T) and they are trying to match end-of-horizon

growth with terminal growth (the growth gap is almost equal to zero in most scenarios, no sudden jumps at the terminal).

- Analysts generally model a quick fade of high growth to a steady state by the end of their 5-7 year forecast. In the sample, companies often go from, say, double-digit growth in early years to low-single-digit by year 5 or 6. The growth gap between last forecast year and terminal is usually small ($<\pm 5\%$), implying analysts try to enter the terminal phase at a natural trajectory. A few exceptions do exist - project-driven surges, for example - but these are handled by imposing a sharp cut to terminal growth implicitly valuing those projects as finite.
- Perhaps the most interesting finding is how analysts deal with changes in risk factors. They do not mechanically update discount rates with each fluctuation in market yields or even each new data point about volatility. Instead, they filter, treating much of the short-term market signal as noise. This is evidenced by the disconnect between soaring local bond yields and unchanged risk-free assumptions in valuations, and by the relatively constant WACCs through very different macro climates. Thus analysts use the long-term averages or subjective opinions while choosing discount rates, and do not update them when there is no specific need. The rational Bayesian updating under uncertainty is consistent with this conclusion.
- Such a conservative and stable attitude towards inputs is probably a good way to be robust against erratic market conditions, but it also can lead to systematic biases. To illustrate this, imagine that Kazakhstan entered some kind of sustained period of structurally lower inflation and interest rates. Analysts would likely be late to cut discount rates, undervaluing the stocks until a new consensus emerged. Of course, in the event of a bubble of overly cheap capital, analysts in this market would have been less likely to drop rates and overvalue—which might be a good thing. Again, the balance between following models vs. using judgement does seem to fall more to judgement here, arguably appropriately for an emerging market with less reliable market signals.

To sum up, the fundamental valuation in the emerging market sample follows the broad patterns identified in global research—discount rates and growth expectations are key drivers of

valuations—but the way analysts choose those inputs reflects their need to navigate an uncertain environment. They effectively apply a heavy dose of common sense and caution: assume things will in the long run look like the rest of the world (hence modest growth and a USD-based risk-free), demand a high return to compensate for myriad risks (hence a high WACC), and don't get swayed by every gyrations in markets or economy (hence slow updating). This leads to valuations that might seem pessimistic on paper—e.g., using a 15% discount rate even when interest rates are low—yet have held up reasonably in practice: many target prices in the sample were in fact not far off market prices subsequently, indicating the market probably shares similar required return views.

For practitioners, the findings underline the importance of stress-testing valuations under various discount rate scenarios—what if country risk truly abates, or conversely, what if it spikes further? Also, analysts and investors should realize that even with these high discount rates, assumptions about terminal value remain a key source of sensitivity: small changes in long-run growth or required return can swing value significantly, given ~55% of value comes from the terminal. Thus, continuing to refine frameworks for estimating sustainable growth (perhaps linking it to long-term real GDP of trading partners, etc.) would add rigor. Finally, this study reinforces that, although textbook models are a starting point, the art of valuation in emerging markets consists of knowing when to deviate from model inputs. The Kazakhstan analysts clearly exercise this art, knowing full well that the actual cost of capital and growth potential are not reflected by today's market metrics. If anything, they provide an interesting example of how informed interpolation between global benchmarks and local realities takes place, a habit that is likely widespread across most emerging markets."

References:

- Bruner, R. F., Eades, K. M., Harris, R. S., & Higgins, R. C. (1998). Best practices in estimating the cost of capital: Survey and synthesis. *Financial Practice and Education*, 8(1), 13–28.*
[ResearchGate+1](#)
- Damodaran, A. (2012). *Investment valuation: Tools and techniques for determining the value of any asset* (3rd ed.). John Wiley & Sons.
https://suhaconsulting.com/wp-content/uploads/2018/09/investment-valuation-3rd-edition.pdf?utm_source=chatgpt.com
- Damodaran, A. (2015). *Country risk: Determinants, measures, and implications – The 2015 edition*. Working paper, New York University Stern School of Business.
<https://doi.org/10.2139/ssrn.2630871>
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2630871&utm_source=chatgpt.com
- Décaire, P. H., & Graham, J. R. (2024). *Valuation fundamentals*. HKU Jockey Club Enterprise Sustainability Global Research Institute – Archive.
<https://doi.org/10.2139/ssrn.4951338>
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4951338&utm_source=chatgpt.com
- Dechow, P. M., & Sloan, R. G. (1997). Returns to contrarian investment strategies: Tests of naive expectations hypotheses. *Journal of Financial Economics*, 43(1), 3–27.
[https://doi.org/10.1016/S0304-405X\(96\)00887-0](https://doi.org/10.1016/S0304-405X(96)00887-0)
https://ideas.repec.org/a/eee/jfinec/v43y1997i1p3-27.html?utm_source=chatgpt.com
- Kaplan, S. N., & Ruback, R. S. (1995). The valuation of cash flow forecasts: An empirical analysis. *Journal of Finance*, 50(4), 1059–1093.
<https://doi.org/10.1111/j.1540-6261.1995.tb04050.x>
https://www.jstor.org/stable/2329344?utm_source=chatgpt.com
- Pereiro, L. E. (2006). The practice of investment valuation in emerging markets: Evidence from Argentina. *Journal of Multinational Financial Management*, 16(2), 160–183.
<https://doi.org/10.1016/j.mulfin.2005.07.002>
https://www.researchgate.net/publication/222614902_The_Practice_of_Investment_Valuation_in_Emerging_Markets_Evidence_from_Argentina?utm_source=chatgpt.com
- Richardson, S. A., Sloan, R. G., Soliman, M. T., & Tuna, I. (2006). The implications of accounting distortions and growth for accruals and profitability. *The Accounting Review*, 81(3), 713–743.
<https://doi.org/10.2308/accr.2006.81.3.713>
https://www.jstor.org/stable/pdf/4093111.pdf?utm_source=chatgpt.com

Appendix

```
from google.colab import drive
drive.mount('/content/drive')

import pandas as pd

file_path = '/content/drive/MyDrive/Data_Thesis_final copy.xlsx'
df = pd.read_excel(file_path)

discount_col = 'WACC'

discount_rates = df[discount_col].astype(str).str.replace(',', '.', regex=False)
discount_rates = pd.to_numeric(discount_rates, errors='coerce').dropna()

import matplotlib.pyplot as plt

plt.figure(figsize=(10, 6))
plt.hist(discount_rates, bins=10)

plt.title('Distribution of Analyst Discount Rates', fontsize=18)
plt.xlabel('Discount Rate (%)', fontsize=14)
plt.ylabel('Number of Valuations', fontsize=14)
plt.grid(True, linestyle='--', alpha=0.5)

plt.tight_layout()
plt.show()

from google.colab import drive
drive.mount('/content/drive')
```

```
import pandas as pd

file_path = '/content/drive/MyDrive/Data_Thesis_final copy.xlsx'
df = pd.read_excel(file_path)

tg_col = 'Term Gr'

terminal_growth = pd.to_numeric(df[tg_col], errors='coerce').dropna()

import matplotlib.pyplot as plt

plt.figure(figsize=(10, 6))
plt.hist(terminal_growth, bins=10)

plt.title('Distribution of Analyst Terminal Growth Rates', fontsize=18)
plt.xlabel('Terminal Growth Rate (%)', fontsize=14)
plt.ylabel('Number of Valuations', fontsize=14)
plt.grid(True, linestyle='--', alpha=0.5)

plt.tight_layout()
plt.show()
```

```

from google.colab import drive

drive.mount('/content/drive')


import pandas as pd
import numpy as np
import matplotlib.pyplot as plt


path = '/content/drive/MyDrive/Data_Thesis_final copy.xlsx'
df = pd.read_excel(path)


print("Columns:", df.columns.tolist())


cf_cols_to_convert = [f'CF{i}\u200b' for i in range(1, 8)]
if 'Base (CF0\u200b)' in df.columns:
    cf_cols_to_convert.append('Base (CF0\u200b)')


for col in cf_cols_to_convert:
    if col in df.columns:
        df[col] = pd.to_numeric(df[col], errors='coerce')


df['WACC'] = pd.to_numeric(df['WACC'], errors='coerce')
df['Term Gr'] = pd.to_numeric(df['Term Gr'], errors='coerce')


def get_horizon(row):
    for i in range(6, 0, -1):
        val = row[f'CF{i}\u200b']
        if pd.notna(val):
            return i
    return np.nan

```

```
df['horizon_years'] = df.apply(get_horizon, axis=1)
```

```
def pv_terms(row):
```

```
    r = row['WACC'] / 100.0
```

```
    g = row['Term Gr'] / 100.0
```

```
    H = row['horizon_years']
```

```
    if pd.isna(r) or pd.isna(g) or pd.isna(H):
```

```
        return pd.Series({'pv_discrete': np.nan, 'pv_terminal': np.nan})
```

```
    H = int(H)
```

```
    pv_disc = 0.0
```

```
    for t in range(1, H+1):
```

```
        cf_t = row[f'CF {t}\u200b']
```

```
        if pd.isna(cf_t):
```

```
            continue
```

```
        pv_disc += cf_t / (1 + r)**t
```

```
    cf_H = row[f'CF {H}\u200b']
```

```
    if pd.isna(cf_H) or r <= g:
```

```
        return pd.Series({'pv_discrete': pv_disc, 'pv_terminal': np.nan})
```

```
    tv_H = cf_H * (1 + g) / (r - g)
```

```
    pv_tv = tv_H / (1 + r)**H
```

```
    return pd.Series({'pv_discrete': pv_disc, 'pv_terminal': pv_tv})
```

```
df[['pv_discrete', 'pv_terminal']] = df.apply(pv_terms, axis=1)
```

```
df['terminal_value_share_pct'] = 100 * df['pv_terminal'] / (df['pv_terminal'] + df['pv_discrete'])
```

```
plot_df = df[['horizon_years', 'terminal_value_share_pct']].dropna()
```

```
plt.figure(figsize=(10, 6))
```

```
plt.scatter(plot_df['horizon_years'], plot_df['terminal_value_share_pct'])
```

```
plt.title('Terminal Value Share vs Forecast Horizon', fontsize=18)
```

```
plt.xlabel('DCF Forecast Horizon (years)', fontsize=14)
```

```
plt.ylabel('Terminal Value Share (%)', fontsize=14)
```

```
plt.grid(True, linestyle='--', alpha=0.5)
```

```
plt.tight_layout()
```

```
plt.show()
```

```
from google.colab import drive
drive.mount('/content/drive')

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

file_path = '/content/drive/MyDrive/Data_Thesis_final copy.xlsx'

main = pd.read_excel(file_path, sheet_name=0)

gdp_raw = pd.read_excel(file_path, sheet_name='Kazakhstan GDP')
gdp_raw = gdp_raw.dropna(how='all')
gdp = pd.DataFrame({
    'Year': pd.to_numeric(gdp_raw.iloc[:, 0], errors='coerce').astype(int),
    'Real_GDP_Growth': pd.to_numeric(gdp_raw.iloc[:, 1], errors='coerce')
}).dropna()

infl_raw = pd.read_excel(file_path, sheet_name='Kazakhstan Inflation')
infl_raw = infl_raw.dropna(how='all')
infl = pd.DataFrame({
    'Year': pd.to_numeric(infl_raw.iloc[:, 0], errors='coerce').astype(int),
    'Inflation_Rate': pd.to_numeric(infl_raw.iloc[:, 1], errors='coerce')
}).dropna()

main['Year'] = pd.to_datetime(main['Date'], format='%b \'%y', errors='coerce').dt.year

merged = (
    main.merge(gdp, on='Year', how='left')
    .merge(infl, on='Year', how='left')
```

)

```
term_col = 'Term Gr'
```

```
merged[term_col] = pd.to_numeric(merged[term_col], errors='coerce')
```

```
def fit_line(x, y):
```

```
    a, b = np.polyfit(x, y, 1)
```

```
    x_line = np.linspace(x.min(), x.max(), 100)
```

```
    y_line = a * x_line + b
```

```
    return x_line, y_line
```

```
gdp_df = merged[['Real_GDP_Growth', term_col]].dropna()
```

```
x_gdp = gdp_df['Real_GDP_Growth'].values
```

```
y_tg = gdp_df[term_col].values
```

```
x_line, y_line = fit_line(x_gdp, y_tg)
```

```
plt.figure(figsize=(6, 4))
```

```
plt.scatter(x_gdp, y_tg)
```

```
plt.plot(x_line, y_line)
```

```
plt.title('Terminal Growth vs Real GDP Growth')
```

```
plt.xlabel('Real GDP Growth (%)')
```

```
plt.ylabel('Terminal Growth Rate (%)')
```

```
plt.grid(True, linestyle='--', alpha=0.5)
```

```
plt.tight_layout()
```

```
plt.show()
```

```
infl_df = merged[['Inflation_Rate', term_col]].dropna()
```

```
x_inf = infl_df['Inflation_Rate'].values
```

```
y_tg2 = infl_df[term_col].values
```

```
x_line2, y_line2 = fit_line(x_inf, y_tg2)
```

```
plt.figure(figsize=(6, 4))
```

```
plt.scatter(x_inf, y_tg2)
```

```
plt.plot(x_line2, y_line2)
```

```
plt.title('Terminal Growth vs Inflation')
```

```
plt.xlabel('Inflation Rate (%)')
```

```
plt.ylabel('Terminal Growth Rate (%)')
```

```
plt.grid(True, linestyle='--', alpha=0.5)
```

```
plt.tight_layout()
```

```
plt.show()
```

```

from google.colab import drive

drive.mount('/content/drive')


import pandas as pd
import numpy as np
import matplotlib.pyplot as plt


file_path = '/content/drive/MyDrive/Data_Thesis_final copy.xlsx'
df = pd.read_excel(file_path, sheet_name=0)


print("Columns:", df.columns.tolist())


cf_cols_to_convert = [f'CF{i}\u200b' for i in range(1, 8)]
other_cols_to_convert = ['WACC', 'Term Gr']


for col in cf_cols_to_convert + other_cols_to_convert:
    if col in df.columns:
        df[col] = pd.to_numeric(df[col], errors='coerce')


def get_horizon(row):
    for i in range(7, 0, -1):
        if pd.notna(row[f'CF{i}\u200b']):
            return i
    return np.nan


df['H'] = df.apply(get_horizon, axis=1)


def pv_blocks(row):
    r = row['WACC'] / 100.0
    g = row['Term Gr'] / 100.0

```

```
H = row['H']
```

```
if pd.isna(r) or pd.isna(g) or pd.isna(H) or r <= g:
```

```
    return pd.Series({'pv_1_3': np.nan, 'pv_4_H': np.nan, 'pv_term': np.nan})
```

```
H = int(H)
```

```
pv_1_3 = 0.0
```

```
for t in range(1, min(3, H) + 1):
```

```
    cf_t = row[f'CF {t}\u200b']
```

```
    if pd.notna(cf_t):
```

```
        pv_1_3 += cf_t / (1 + r) ** t
```

```
pv_4_H = 0.0
```

```
if H > 3:
```

```
    for t in range(4, H + 1):
```

```
        cf_t = row[f'CF {t}\u200b']
```

```
        if pd.notna(cf_t):
```

```
            pv_4_H += cf_t / (1 + r) ** t
```

```
cf_H = row[f'CF {H}\u200b']
```

```
if pd.isna(cf_H) or r <= g:
```

```
    return pd.Series({'pv_1_3': pv_1_3, 'pv_4_H': pv_4_H, 'pv_term': np.nan})
```

```
tv_H = cf_H * (1 + g) / (r - g)
```

```
pv_term = tv_H / (1 + r) ** H
```

```
return pd.Series({'pv_1_3': pv_1_3, 'pv_4_H': pv_4_H, 'pv_term': pv_term})
```

```
df[['pv_1_3', 'pv_4_H', 'pv_term']] = df.apply(pv_blocks, axis=1)
```

```
df['pv_total'] = df[['pv_1_3', 'pv_4_H', 'pv_term']].sum(axis=1)
```

```
df['sh_1_3'] = df['pv_1_3'] / df['pv_total']
```

```
df['sh_4_H'] = df['pv_4_H'] / df['pv_total']
```

```
df['sh_term'] = df['pv_term'] / df['pv_total']
```

```
shares_df = df.dropna(subset=['H', 'sh_1_3', 'sh_4_H', 'sh_term']).copy()
```

```
group_means = shares_df.groupby('H')[['sh_1_3', 'sh_4_H', 'sh_term']].mean()
```

```
horizons_to_keep = [h for h in [4, 5, 6, 7] if h in group_means.index]
```

```
group_means = group_means.loc[horizons_to_keep]
```

```
overall = shares_df[['sh_1_3', 'sh_4_H', 'sh_term']].mean().to_frame('All').T
```

```
plot_data = pd.concat([group_means, overall], axis=0)
```

```
index_labels = [f'{int(h)}-yr' for h in horizons_to_keep] + ["All"]
```

```
plot_data.index = index_labels
```

```
print("\nAverage PV shares by horizon:")
```

```
print(plot_data)
```

```
labels = plot_data.index
```

```
x = np.arange(len(labels))
```

```
y1 = plot_data['sh_1_3'].values
```

```
y2 = plot_data['sh_4_H'].values
```

```
y3 = plot_data['sh_term'].values
```

```
plt.figure(figsize=(8, 4))
```

```
bottom0 = np.zeros(len(labels))
plt.bar(x, y1, bottom=bottom0, label='Years 1–3')

bottom1 = bottom0 + y1
plt.bar(x, y2, bottom=bottom1, label='Years 4–Horizon')

bottom2 = bottom1 + y2
plt.bar(x, y3, bottom=bottom2, label='Terminal Value')

plt.xticks(x, labels)
plt.ylim(0, 1)
plt.ylabel('Share of DCF Present Value')
plt.title('Average DCF Value Contribution by Cash Flow Block and Horizon')
plt.grid(True, axis='y', linestyle='--', alpha=0.5)
plt.legend()
plt.tight_layout()
plt.show()

from google.colab import drive
drive.mount('/content/drive')
```

```

import pandas as pd
import numpy as np
from scipy.stats import pearsonr

file_path = '/content/drive/MyDrive/Data_Thesis_final copy.xlsx'

main = pd.read_excel(file_path, sheet_name=0)

gdp_raw = pd.read_excel(file_path, sheet_name='Kazakhstan GDP')
infl_raw = pd.read_excel(file_path, sheet_name='Kazakhstan Inflation')

gdp = gdp_raw.dropna(how='all')
gdp = pd.DataFrame({
    'Year': pd.to_numeric(gdp.iloc[:, 0], errors='coerce').astype(int),
    'Real_GDP_Growth': pd.to_numeric(gdp.iloc[:, 1], errors='coerce')
}).dropna()

infl = infl_raw.dropna(how='all')
infl = pd.DataFrame({
    'Year': pd.to_numeric(infl.iloc[:, 0], errors='coerce').astype(int),
    'Inflation_Rate': pd.to_numeric(infl.iloc[:, 1], errors='coerce')
}).dropna()

main['Year'] = pd.to_datetime(main['Date'], format='%b %y', errors='coerce').dt.year

merged = (
    main.merge(gdp, on='Year', how='left')
    .merge(infl, on='Year', how='left')
)

```

```

term_col = 'Term Gr'

merged[term_col] = pd.to_numeric(merged[term_col], errors='coerce')

rows = []

def add_corr_row(label, x_col):
    data = merged[[term_col, x_col]].dropna().copy()
    if len(data) < 3:
        return

    col1 = data[term_col].astype(float)
    col2 = data[x_col].astype(float)

    r, p = pearsonr(col1, col2)

    if abs(r) < 0.1:
        interp = f"Essentially zero (p \u2248 {p:.2f})"
    elif r > 0:
        interp = f"Positive but modest (p \u2248 {p:.2f})"
    else:
        interp = f"Negative but modest (p \u2248 {p:.2f})"

    rows.append({
        "Correlation": label,
        "Pearson \u03C11": f"{r:+.2f}",
        "Interpretation": interp
    })

add_corr_row(r"$G_T$ vs. Real GDP Growth", "Real_GDP_Growth")

```

```
add_corr_row(r"$G_T$ vs. Inflation", "Inflation_Rate")
```

```
table6 = pd.DataFrame(rows,  
    columns=["Correlation", "Pearson \u03C1", "Interpretation"])
```

```
print("\nTable 6. Correlation of Terminal Growth with Macro Factors\n")
```

```
print(table6.to_string(index=False))
```

```

from google.colab import drive
drive.mount('/content/drive')

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

file_path = '/content/drive/MyDrive/Data_Thesis_final copy.xlsx'
df = pd.read_excel(file_path, sheet_name=0)

print("Columns:", df.columns.tolist())

cf_cols_to_convert = [f'CF {i}\u200b' for i in range(1, 8)]
for col in cf_cols_to_convert:
    if col in df.columns:
        df[col] = pd.to_numeric(df[col], errors='coerce')

df['Term Gr'] = pd.to_numeric(df['Term Gr'], errors='coerce')

def get_horizon(row):
    for i in range(7, 0, -1):
        if pd.notna(row[f'CF {i}\u200b']):
            return i
    return np.nan

df['H'] = df.apply(get_horizon, axis=1)

def last_explicit_growth(row):
    H = row['H']
    if pd.isna(H) or H < 2:

```

```

        return np.nan
    H = int(H)
    cf_H = row[f'CF{H}\u200b']
    cf_prev = row[f'CF{H-1}\u200b']
    if pd.isna(cf_H) or pd.isna(cf_prev) or cf_prev == 0:
        return np.nan
    return cf_H / cf_prev - 1.0

df['g_last_dec'] = df.apply(last_explicit_growth, axis=1)

df['g_last_pct'] = df['g_last_dec'] * 100.0
df['growth_gap_pct'] = df['g_last_pct'] - df['Term Gr']

print("\nGrowth gap summary (percentage points):")
print(df['growth_gap_pct'].describe())

gap = df['growth_gap_pct'].dropna()

plt.figure(figsize=(10, 6))
plt.hist(gap, bins=20)

plt.title('Distribution of "Growth Gap"\n(Last Explicit Year Growth \u2212 Terminal Growth)', fontsize=16)
plt.xlabel('Growth Gap (percentage points)', fontsize=14)
plt.ylabel('Number of Valuations', fontsize=14)
plt.grid(True, linestyle='--', alpha=0.5)

plt.tight_layout()
plt.show()

```

```

import math

from typing import Dict, Optional


import numpy as np
import pandas as pd
import statsmodels.api as sm
import matplotlib.pyplot as plt


from google.colab import drive
drive.mount('/content/drive')


data_path = "/content/drive/MyDrive/Data_Thesis_final.xlsx"
dcf_df = pd.read_excel(data_path)


def compute_growth_rates_from_fcf(row: pd.Series,
                                   cf0_col: str = "cf0"
                                   ) -> Dict[str, Optional[float]]:

    cf0 = row.get(cf0_col, None)
    f1 = row.get("fcf_t1", None)
    f2 = row.get("fcf_t2", None)
    f3 = row.get("fcf_t3", None)


def safe_growth(a, b):
    if a is None or b is None:
        return None

    try:
        if a == 0:
            return None

        ratio = b / a

```

```

    if ratio <= 0:
        return None
    return ratio - 1.0
except ZeroDivisionError:
    return None

```

```

g1 = safe_growth(cf0, f1)
g2 = safe_growth(f1, f2)
g3 = safe_growth(f2, f3)

```

```

return {"g1": g1, "g2": g2, "g3": g3}

```

```

def build_cs_dataset(dcf_df: pd.DataFrame,
                    cf0_col: str = "cf0",
                    value_col: str = "equity_value",
                    rho: float = 0.96) -> pd.DataFrame:
    df = dcf_df.copy()

    g1_list, g2_list, g3_list = [], [], []
    for _, row in df.iterrows():
        gr = compute_growth_rates_from_fcf(row, cf0_col=cf0_col)
        g1_list.append(gr["g1"])
        g2_list.append(gr["g2"])
        g3_list.append(gr["g3"])
    df["g1"] = g1_list
    df["g2"] = g2_list
    df["g3"] = g3_list

    def safe_log_ratio(v, cf0):

```

```
if v is None or cf0 is None:
```

```
    return None
```

```
try:
```

```
    if cf0 <= 0 or v <= 0:
```

```
        return None
```

```
    return math.log(v / cf0)
```

```
except (ValueError, ZeroDivisionError):
```

```
    return None
```

```
df["log_valuation_ratio"] = [
```

```
    safe_log_ratio(row.get(value_col), row.get(cf0_col))
```

```
    for _, row in df.iterrows()
```

```
]
```

```
df["gT_dec"] = df["terminal_growth_pct"] / 100.0
```

```
df["r_dec"] = df["discount_rate_pct"] / 100.0
```

```
def cs_components_row(row):
```

```
    H = row.get("horizon_years", 0)
```

```
    g1 = row.get("g1", None)
```

```
    g2 = row.get("g2", None)
```

```
    g3 = row.get("g3", None)
```

```
    gT = row.get("gT_dec", None)
```

```
    r = row.get("r_dec", None)
```

```
    if r is None:
```

```
        return (None, None, None, None)
```

```
    short = None if g1 is None else math.log(1 + g1)
```

```

medium = 0.0

valid_med = False

if g2 is not None:
    medium += rho * math.log(1 + g2)
    valid_med = True

if g3 is not None:
    medium += (rho**2) * math.log(1 + g3)
    valid_med = True

if not valid_med:
    medium = None


long_term = 0.0
valid_long = False


for i in range(4, H + 1):
    gi = None
    if i <= 10:
        f_i = row.get(f'fcf_t{i}', None)
        f_prev = row.get(f'fcf_t{i-1}', None)
        if f_i is not None and f_prev is not None and f_prev != 0:
            gi = f_i / f_prev - 1.0
    if gi is not None:
        long_term += (rho ** (i - 1)) * math.log(1 + gi)
        valid_long = True


if (gT is not None) and (H is not None):
    try:
        long_term += (rho**H) / (1 - rho) * math.log(1 + gT)
        valid_long = True
    except ValueError:

```

```
pass
```

```
if not valid_long:
```

```
    long_term = None
```

```
discount_term = None
```

```
try:
```

```
    discount_term = - (1.0 / (1.0 - rho)) * math.log(1 + r)
```

```
except ValueError:
```

```
    discount_term = None
```

```
return (short, medium, long_term, discount_term)
```

```
cs_short, cs_medium, cs_long, cs_disc = [], [], [], []
```

```
for _, row in df.iterrows():
```

```
    s, m, l, d = cs_components_row(row)
```

```
    cs_short.append(s)
```

```
    cs_medium.append(m)
```

```
    cs_long.append(l)
```

```
    cs_disc.append(d)
```

```
df["cs_short"] = cs_short
```

```
df["cs_medium"] = cs_medium
```

```
df["cs_long"] = cs_long
```

```
df["cs_discount"] = cs_disc
```

```
return df
```

```
def run_cs_regressions(cs_df: pd.DataFrame):
```

```
df = cs_df.dropna(
    subset=["log_valuation_ratio",
           "cs_short", "cs_medium", "cs_long", "cs_discount"]
).copy()
```

```
results = {}
```

```
def ols(y, X, label):
    X = sm.add_constant(X)
    model = sm.OLS(y, X)
    res = model.fit(cov_type="HC3")
    results[label] = res
    print(f"\n===== Regression: {label} =====")
    print(res.summary())
    return res
```

```
ols(df["log_valuation_ratio"], df["cs_short"], "short_only")
ols(df["log_valuation_ratio"], df["cs_medium"], "medium_only")
ols(df["log_valuation_ratio"], df["cs_long"], "long_only")
ols(df["log_valuation_ratio"], df["cs_discount"], "discount_only")
```

```
ols(
    df["log_valuation_ratio"],
    df[["cs_short", "cs_medium", "cs_long", "cs_discount"]],
    "all_components"
)
```

```
return results
```

```
def basic_summary_statistics(cs_df: pd.DataFrame):

    cols = [

        "log_valuation_ratio",

        "g1", "g2", "g3",

        "terminal_growth_pct",

        "discount_rate_pct"

    ]

    print("\n=== Summary statistics (selected variables) ===")

    print(cs_df[cols].describe())
```

```
def plot_time_series(df: pd.DataFrame,

                    date_col: str,

                    value_col: str,

                    title: str):

    tmp = df.copy()

    tmp[date_col] = pd.to_datetime(tmp[date_col])

    tmp = tmp.dropna(subset=[date_col, value_col])

    tmp = tmp.sort_values(date_col)
```

```
plt.figure()

plt.plot(tmp[date_col], tmp[value_col])

plt.xlabel("Date")

plt.ylabel(value_col)

plt.title(title)

plt.grid(True)

plt.tight_layout()

plt.show()
```

```

if __name__ == "__main__":

    dcf_df.columns = dcf_df.columns.str.strip().str.replace('\u200b', '')
    print("\nColumns after stripping whitespace and unicode chars:", dcf_df.columns.tolist())

    rename_map = {
        "Actual Price": "equity_value",
        "WACC": "discount_rate_pct",
        "Term Gr": "terminal_growth_pct",
        "Base (CF0)": "cf0",
    }

    for i in range(0, 7):
        rename_map[f"CF {i}"] = f"fcf_t{i}"

    dcf_df = dcf_df.rename(columns=rename_map)

    print("Columns after renaming:", dcf_df.columns.tolist())

    numeric_cols = [
        "equity_value", "cf0",
        "fcf_t1", "fcf_t2", "fcf_t3", "fcf_t4", "fcf_t5", "fcf_t6", "fcf_t7",
        "discount_rate_pct", "terminal_growth_pct"
    ]

    for col in numeric_cols:
        if col in dcf_df.columns:
            dcf_df[col] = pd.to_numeric(dcf_df[col], errors='coerce')

    cs_df = build_cs_dataset(
        dcf_df,

```

```
cf0_col="cf0",  
value_col="equity_value",  
rho=0.96  
)  
  
cs_df.to_csv("Data_Thesis_CS_components.csv", index=False)  
print("Campbell–Shiller dataset saved to: Data_Thesis_CS_components.csv")  
  
basic_summary_statistics(cs_df)  
  
cs_results = run_cs_regressions(cs_df)
```

```
from google.colab import drive  
drive.mount('/content/drive')
```

```

import pandas as pd
import numpy as np
import statsmodels.api as sm
from scipy.stats import pearsonr, spearmanr

file_path = '/content/drive/MyDrive/Data_Thesis_final copy.xlsx'

main = pd.read_excel(file_path, sheet_name=0)

gdp_raw = pd.read_excel(file_path, sheet_name='Kazakhstan GDP')
infl_raw = pd.read_excel(file_path, sheet_name='Kazakhstan Inflation')

disc = pd.read_excel(file_path, sheet_name='Discount Rate Regression')

bond_raw = pd.read_excel(file_path, sheet_name='Kazakhstan 10 year bond')

print("Main columns:", main.columns.tolist())
print("Discount-rate columns:", disc.columns.tolist())
print("Bond columns:", bond_raw.columns.tolist())

cf_cols_to_convert = [f'CF {i}\u200b' for i in range(1, 8)]
if 'Base (CF0\u200b)' in main.columns:
    cf_cols_to_convert.append('Base (CF0\u200b)')

other_dcf_cols = ['WACC', 'Term Gr']

for col in cf_cols_to_convert + other_dcf_cols:
    if col in main.columns:
        main[col] = pd.to_numeric(main[col], errors='coerce')

```

```

term_col = 'Term Gr'
wacc_col_main = 'WACC'
cf_cols = [f'CF {i}\u200b' for i in range(1, 8)]

def compute_horizon(row):
    for i in range(7, 0, -1):
        if pd.notna(row[f'CF {i}\u200b']):
            return i
    return np.nan

main['H'] = main.apply(compute_horizon, axis=1)

def pv_blocks(row):
    """Return PV of years 1–3, 4–H and terminal leg."""
    r = row[wacc_col_main] / 100.0
    g = row[term_col] / 100.0
    H = row['H']
    if pd.isna(r) or pd.isna(g) or pd.isna(H) or r <= g:
        return pd.Series({'pv_1_3': np.nan, 'pv_4_H': np.nan, 'pv_term': np.nan})
    H = int(H)

    pv_1_3 = 0.0
    for t in range(1, min(3, H) + 1):
        cf_t = row[f'CF {t}\u200b']
        if pd.notna(cf_t):
            pv_1_3 += cf_t / (1 + r) ** t

    pv_4_H = 0.0
    if H > 3:
        for t in range(4, H + 1):

```

```

cf_t = row[f'CF {t}\u200b']

if pd.notna(cf_t):
    pv_4_H += cf_t / (1 + r) ** t

cf_H = row[f'CF {H}\u200b']

if pd.isna(cf_H):
    pv_term = np.nan
else:
    tv_H = cf_H * (1 + g) / (r - g)
    pv_term = tv_H / (1 + r) ** H

return pd.Series({'pv_1_3': pv_1_3, 'pv_4_H': pv_4_H, 'pv_term': pv_term})

main[['pv_1_3', 'pv_4_H', 'pv_term']] = main.apply(pv_blocks, axis=1)
main['pv_total'] = main[['pv_1_3', 'pv_4_H', 'pv_term']].sum(axis=1)
main['sh_1_3'] = main['pv_1_3'] / main['pv_total']
main['sh_4_H'] = main['pv_4_H'] / main['pv_total']
main['sh_term'] = main['pv_term'] / main['pv_total']

def last_year_growth(row):
    """g_H = CF_H / CF_{H-1} - 1 (decimal)."""
    H = row['H']
    if pd.isna(H) or H < 2:
        return np.nan
    H = int(H)
    cf_H = row[f'CF {H}\u200b']
    cf_Hm1 = row[f'CF {H-1}\u200b']
    if pd.isna(cf_H) or pd.isna(cf_Hm1) or cf_Hm1 == 0:
        return np.nan
    return cf_H / cf_Hm1 - 1.0

```

```

main['g_last_dec']    = main.apply(last_year_growth, axis=1)
main['g_last_pct']    = main['g_last_dec'] * 100.0
main['growth_gap_pct'] = main['g_last_pct'] - main[term_col]

```

```

def avg_last2_growth(row):
    """Average of growth between H-2→H-1 and H-1→H (decimal)."""
    H = row['H']
    if pd.isna(H) or H < 3:
        return np.nan
    H = int(H)
    cf_H  = row[f'CF{H}\u200b']
    cf_Hm1 = row[f'CF{H-1}\u200b']
    cf_Hm2 = row[f'CF{H-2}\u200b']
    if any(pd.isna([cf_H, cf_Hm1, cf_Hm2])) or cf_Hm1 == 0 or cf_Hm2 == 0:
        return np.nan
    g1 = cf_Hm1 / cf_Hm2 - 1.0
    g2 = cf_H  / cf_Hm1 - 1.0
    return (g1 + g2) / 2.0

```

```

main['g_last2_avg_dec'] = main.apply(avg_last2_growth, axis=1)
main['g_last2_avg_pct'] = main['g_last2_avg_dec'] * 100.0
main['growth_gap2_pct'] = main['g_last2_avg_pct'] - main[term_col]

```

```

main['Year'] = pd.to_datetime(main['Date'], format='%b %y', errors='coerce').dt.year

```

```

gdp = gdp_raw.dropna(how='all')
gdp = pd.DataFrame({
    'Year': pd.to_numeric(gdp.iloc[:, 0], errors='coerce').astype(int),
    'Real_GDP_Growth': pd.to_numeric(gdp.iloc[:, 1], errors='coerce')
})

```

```
}).dropna()
```

```
infl = infl_raw.dropna(how='all')
```

```
infl = pd.DataFrame({  
    'Year': pd.to_numeric(infl.iloc[:, 0], errors='coerce').astype(int),  
    'Inflation_Rate': pd.to_numeric(infl.iloc[:, 1], errors='coerce')  
}).dropna()
```

```
macro = (gdp.merge(infl, on='Year', how='outer')  
         .sort_values('Year')  
         .reset_index(drop=True))
```

```
macro['GDP_3y_ma'] = macro['Real_GDP_Growth'].rolling(3, min_periods=1).mean()  
macro['Infl_3y_ma'] = macro['Inflation_Rate'].rolling(3, min_periods=1).mean()
```

```
main_macro = main.merge(macro, on='Year', how='left')
```

```
def describe_series(name, s):  
    print(f"\n=== {name} ===")  
    print(s.describe())
```

```
def winsorize(s, lower=0.01, upper=0.99):  
    lo, hi = s.quantile([lower, upper])  
    return s.clip(lo, hi)
```

```
from collections import OrderedDict
```

```
def corr_summary(df, x, y, label):  
    tmp = df[[x, y]].dropna()  
    if len(tmp) < 3:
```

```

        print(f"{label}: too few observations")

        return

    tmp[x] = tmp[x].astype(float)
    tmp[y] = tmp[y].astype(float)

    r_p, p_p = pearsonr(tmp[x], tmp[y])
    r_s, p_s = spearmanr(tmp[x], tmp[y])
    print(f"{label}:")
    print(f"  n = {len(tmp)}")
    print(f"  Pearson r = {r_p:.3f} (p = {p_p:.3f})")
    print(f"  Spearman rho = {r_s:.3f} (p = {p_s:.3f})")

def ols_robust(df, y_col, x_cols, label, cov_type="HC3"):
    tmp = df[[y_col] + x_cols].dropna()
    if len(tmp) < len(x_cols) + 5:
        print(f"{label}: too few observations for regression")
        return None

    for col in [y_col] + x_cols:
        tmp[col] = tmp[col].astype(float)

    y = tmp[y_col]
    X = sm.add_constant(tmp[x_cols])
    model = sm.OLS(y, X).fit(cov_type=cov_type)
    print(f"\n=== {label} ===")
    print(model.summary())
    return model

print("\n\nR1. Outliers & Winsorization")

wacc = main[wacc_col_main].dropna()

```

```

gt = main[term_col].dropna()
gap1 = main['growth_gap_pct'].dropna()

describe_series("WACC (baseline)", wacc)
describe_series("Terminal growth (baseline)", gt)
describe_series("Growth gap (last-year, baseline)", gap1)

describe_series("WACC (winsor 1/99)", winsorize(wacc, 0.01, 0.99))
describe_series("WACC (winsor 2.5/97.5)", winsorize(wacc, 0.025, 0.975))

describe_series("GT (winsor 1/99)", winsorize(gt, 0.01, 0.99))
describe_series("Growth gap (winsor 2.5/97.5)", winsorize(gap1, 0.025, 0.975))

q_low, q_high = gap1.quantile([0.05, 0.95])
gap_trim = gap1[(gap1 >= q_low) & (gap1 <= q_high)]
describe_series("Growth gap (trimmed 5% tails)", gap_trim)

print("\n\nR2. Firm-level collapse")

cols_for_firm = ['Firm', wacc_col_main, term_col, 'sh_term', 'growth_gap_pct']
firm_level = (main[cols_for_firm]
              .dropna(subset=['Firm'])
              .groupby('Firm', as_index=False)
              .mean())

describe_series("Firm-level WACC", firm_level[wacc_col_main])
describe_series("Firm-level GT", firm_level[term_col])
describe_series("Firm-level terminal share", firm_level['sh_term'])

print("\nMeans: report vs firm-level")

```

```

print(pd.DataFrame({
    'report_level': [
        main[wacc_col_main].mean(),
        main[term_col].mean(),
        main['sh_term'].mean()
    ],
    'firm_level': [
        firm_level[wacc_col_main].mean(),
        firm_level[term_col].mean(),
        firm_level['sh_term'].mean()
    ]
}, index=["WACC", 'GT', 'TV share']))

print("\n\n R3. GT vs macro robustness ")

corr_summary(main_macro, 'Real_GDP_Growth', term_col,
    "GT vs Real GDP (current year)")

corr_summary(main_macro, 'Inflation_Rate', term_col,
    "GT vs Inflation (current year)")

corr_summary(main_macro, 'GDP_3y_ma', term_col,
    "GT vs 3-year avg Real GDP")

corr_summary(main_macro, 'Infl_3y_ma', term_col,
    "GT vs 3-year avg Inflation")

ols_robust(main_macro, term_col,
    ['Real_GDP_Growth', 'Inflation_Rate'],
    "GT regression on current GDP and inflation")

ols_robust(main_macro, term_col,

```

```

['GDP_3y_ma', 'Infl_3y_ma'],
"GT regression on 3-year avg GDP and inflation")

print("\n\nR4. WACC determinants & updating ")

bond = bond_raw.dropna(how='all').copy()

if 'Year' in bond.columns:
    bond['Year'] = bond['Year'].astype(int)
elif 'Date' in bond.columns:
    bond['Year'] = pd.to_datetime(bond['Date'], format='%b %y', errors='coerce').dt.year
else:
    bond['Year'] = pd.to_numeric(bond.iloc[:, 0], errors='coerce').astype(int)

if 'yield_10y_pct' not in bond.columns:
    if 'Price' in bond.columns:
        bond['yield_10y_pct'] = pd.to_numeric(bond['Price'], errors='coerce')
    else:
        bond['yield_10y_pct'] = pd.to_numeric(bond.iloc[:, 1], errors='coerce')
bond = bond.dropna(subset=['Year', 'yield_10y_pct'])

disc = disc.copy()
disc['beta_erp_pct'] = disc['beta_levered'] * disc['equity_risk_premium_pct']
disc = disc.merge(main[['Report Source', 'Firm', 'Date', 'Year']],
                  left_on='report_id', right_on='Report Source', how='left')
disc = disc.merge(bond[['Year', 'yield_10y_pct']], on='Year', how='left')

ols_robust(disc, 'wacc_pct',
           ['risk_free_rate_pct', 'beta_erp_pct'],
           "WACC regression on RF and beta*ERP")

```

```

ols_robust(disc, 'wacc_pct',
           ['risk_free_rate_pct', 'beta_erp_pct', 'yield_10y_pct'],
           "WACC regression incl. 10-year KZT yield")

disc['Date'] = pd.to_datetime(disc['Date'], format='%b %y', errors='coerce')
disc = disc.sort_values(['Firm', 'Date'])

disc['d_wacc'] = disc.groupby('Firm')['wacc_pct'].diff()
disc['d_rf'] = disc.groupby('Firm')['risk_free_rate_pct'].diff()
disc['d_betaerp'] = disc.groupby('Firm')['beta_erp_pct'].diff()

ols_robust(disc, 'd_wacc',
           ['d_rf', 'd_betaerp'],
           "Change in WACC on changes in RF and beta*ERP")

print("\n\n R5. Alternative growth-gap definitions")

gap1 = main['growth_gap_pct'].dropna()
gap2 = main['growth_gap2_pct'].dropna()

describe_series("Growth gap 1 (last-year growth - GT)", gap1)
describe_series("Growth gap 2 (avg of last 2 years - GT)", gap2)

q1_low, q1_high = gap1.quantile([0.05, 0.95])
q2_low, q2_high = gap2.quantile([0.05, 0.95])
gap1_trim = gap1[(gap1 >= q1_low) & (gap1 <= q1_high)]
gap2_trim = gap2[(gap2 >= q2_low) & (gap2 <= q2_high)]

describe_series("Growth gap 1 (trimmed 5%)", gap1_trim)

```

```
describe_series("Growth gap 2 (trimmed 5%)", gap2_trim)
```

```
print("\nDone")
```