NPL
National Physical Laboratory

NPL Report Number IEA 35

Optimal Testing in the Economy

Mike King, Zahrah Qureshi, Dharani Reddy Katanguru

JULY 2026

National Physical Laboratory (NPL)

Optimal Testing in the Economy

Mike King, Zahrah Qureshi, Dharani Reddy Katanguru

We are the UK's National Metrology Institute (NMI), a world-leading centre of excellence that provides cutting-edge measurement science, engineering and technology that underpins prosperity and quality of life in the UK.

Abstract

Measurement is a fundamental part of an economic infrastructure, supporting the reliability and consistency of information used across production, trade, and regulation. The UK's National Measurement System (NMS) provides the measurement infrastructure that supports accurate, reliable, and internationally comparable measurements throughout the economy. Much of the economic value generated by the NMS comes through Delivering the Measurement Infrastructure (traceability benefits). King and Nayak (2025) develop a framework for these traceability benefits bringing together a number of interconnected and technically complex elements. As a result, the framework can be challenging for non-specialist audiences to engage with. This report focuses on unpacking the first component of that framework, namely, the macroeconomic model of conformance testing. It provides a technical exposition that complements the underlying mathematics with economic intuition. The model characterises the equilibrium level of conformance testing in the economy and enables analysis of how changes in underlying parameters alter the effectiveness of testing. To support interpretation, the framework is operationalised through a calculator that enables exploration of equilibrium outcomes and comparative statics, highlighting the critical role of high-quality, traceable measurement in sustaining productivity across the economy.

MCR-001, Issue 1, July 2025

ISSN 2633-4194

https://doi.org/10.47120/npl.IEA35

National Physical Laboratory

Hampton Road, Teddington, Middlesex, TW11 0LW

This work was funded by the UK Government's Department for Science, Innovation & Technology through the UK's National Measurement System programmes.

Contents

Executive Summary

By funding measurement science at specialist laboratories, the **UK's National Measurement System (NMS)** provides the infrastructure which ensures the accuracy, reliability, and international comparability of measurements made across the economy. The impact flows through two main economic channels:

1. **Enabling Technological Change** (*Innovation Benefits*):
   Support from the NMS laboratories through collaborative research and development activities help firms to strengthen their capabilities for effective in-house R&D which leads to innovation.
2. **Delivering the Measurement Infrastructure** (*Traceability Benefits*):
   This channel is based around defining the NMS labs as the source of certified reference materials and calibrations, traceable to national standards, but delivered through a network of accredited labs.

Much of the existing economic literature on measurement and standards focuses on 'innovation benefits' through technological change. In contrast, the 'traceability benefits' relate to the more specialised domain of metrology, which has received limited attention from mainstream economics. To address this gap, NPL economists have constructed a framework that draws on literature which identifies measurement as a foundational "infra technology" with economy wide effects.

King and Nayak (2025) builds an economic framework for quantifying the benefit of sustained public investment in the **UK's Measurement Infrastructure** delivered through the NMS programme. While much of this framework can be found in the study, there are additional models that fully make up the framework for the 'traceability benefits'. The framework is seen to comprise the following five components:

MCR-001, Issue 1, July 2025

1. **Gravity Model:** It explains the impact of distance from an NMI and the uptake of quality calibrations.
2. **Calibration and Fanout Model (work-in-progress):** It explains the distribution and transportation of information through the instruments across the tiers of a calibration chain.
3. **Basic and Precise Calibrations Model:** It explains how precise calibrations supplied by an NMI reduces systematic measurement error and improves testing decisions.
4. **Type I and Type II Errors Model:** It describes how measurement uncertainty in conformance testing leads to decision errors, false positives (Type I errors), and false negatives (Type II errors).
5. **Optimal Level of Testing Model:** The Optimal Testing Model characterises the equilibrium level of conformance testing in the economy and enables analysis of how changes in underlying parameters alter the effectiveness of testing. Crucially, it provides a platform for quantifying the changes in demand for conformance testing attributable to public funding of NMS laboratories and their provision of high-quality calibrations.

This report is dedicated to explaining and exhibiting the *Optimal Testing Model* found within the much longer original study alongside the other interlinked models listed above. The scale and complexity of the original study somewhat obscured the Optimal Testing Model which is better understood as a standalone model and worthy of its own report.

The Optimal Testing Model is centred on the economically efficient level of conformance testing in an economy where production processes are imperfect and capital equipment can malfunction. The model is built around three key concepts that determine the optimal level of testing: the reliability of production, the detection rate, and the regret rate. Reliability captures the proportion of production generated by functioning capital that produces output conforming to specification. Conformance testing improves reliability by identifying defective output generated by malfunctioning capital equipment and enabling that capital to be reset and restored to normal operation. The detection rate refers to the rate at which defective output is correctly identified and malfunctioning capital is restored, while the regret rate refers to the rate at which conforming output is erroneously rejected in the testing process (a Type I error). Conformance testing also requires dedicated engineers whose wages must be funded from a share of production, giving rise to a form of "reliability tax" on capital owners. As testing intensity increases, however, diminishing marginal returns emerge that is additional rounds of testing continue to improve reliability, but the benefits are gradually offset by rising testing costs and increasing losses from Type I errors, where conforming output is mistakenly scrapped.

The macroeconomic approach taken in this study adapts the Solow model to compute the optimal frequency of conformance testing that maximises consumption per worker (where the economy's capital intensity and the output per worker are also maximised) – A concept analogous to the 'Golden Rule Savings Rate'. The Golden Rule identifies the optimal equilibrium level of testing for a given set of parameters. At the optimal frequency, the marginal increase in benefits from additional testing exactly offset the loss from additional costs. Comparative statics is then used to examine how this equilibrium shifts when a basic parameter changes.

The model retains much of the elegance of the original Solow model and its equilibrium reduces to a pair of simultaneous equations with two unknowns: Detection rate (*rate at which defective output is detected, and the malfunctioning part of capital is reset*) and the regret rate (*rate at which good output gets scrapped erroneously*). Using numerical methods to solve for these unknowns then allows all other components of the model to be found. A **key finding** from the model is the **elasticity of the reliability of production with respect to the testing frequency**. The model finds that by doubling the frequency of testing, the reliability of production increases by roughly **2.4%**.

This report also accompanies a calculator that operationalises the framework by enabling the computation of the baseline equilibrium for the economy under existing measurement conditions and performs comparative statics by varying key parameters. For instance, small changes in the probability of Type I errors reduce the optimal frequency of testing, lower detection rates, and diminish output, highlighting the role of high quality, traceable measurements delivered by the NMS labs. The model is derived from estimates that are based on data covering 2015-2019. An updated report is currently in preparation and is expected to be published in September 2026. It will present revised estimates based on data for 2021-2023 and provide a more detailed theoretical explanation of its components.[1]

[1] The link to the calculator can be found in the Annex A Guide for the Calculator

1 Introduction

1.1 Background

The National Measurement System (NMS) provides the technical and organisational infrastructure that ensures measurements made across the UK are accurate, reliable, and internationally recognised. As set out in the Measuring Up to the Competition (1989) white paper, its purpose is to: enable individuals and organisations to make accurate, defensible measurements; and ensure coordination of the UK's measurement system with those of other countries.

This role is distinct from research and development focused on new technologies: measurement science instead underpins the reliability of information generated across both routine economic activity and innovation. All measurement devices from industrial instruments to sensors embedded in automated systems require calibration to verify that their outputs are correct; without this, increased measurement uncertainty leads to a higher incidence of incorrect decisions in production, product verification, and real-time process control. As economies become more reliant on data-driven and autonomous systems, the importance of reliable measurement information continues to grow. Internationally, the NMS supports trust and comparability through participation in global measurement frameworks and key comparisons, reducing non-tariff barriers that can hamper cross-border trade. At the same time, evidence of a strong home-bias[2] in calibration services highlights the national importance of domestic NMS capability, with firms preferring locally accessible, trusted, and traceable measurement services over reliance on foreign providers.

The main way in which measurement supports the broader economy is through the *Testing, Inspections and Certifications* (TIC) sector, which is the backbone of the UK's quality infrastructure, providing third-party assurance that products, processes, and services meet rigorous safety and regulatory standards. While TIC acts as the "verifier," it relies on measurement science for its scientific validity. Measurement science provides the precise standards and traceable calibration needed to ensure that a test, or inspection, conducted

[2] As theorized in King and Renedo (2020).

by a TIC firm is accurate and globally recognised. This system of co-dependence is referred to in its entirety as the UK's National Quality Infrastructure (NQI). It is a system that underpins trust, safety, and reliability in products and services by providing an infrastructure that supports regulatory compliance, facilitates trade, and removes technical barriers by aligning UK practices with international standards. Its five core components are:

1. **Standardisation** – developing and maintaining standards that codify best practice.
2. **Accreditation** – assuring the competence and integrity of organisations providing testing, inspection, and certification.
3. **Metrology** – maintaining accurate reference standards for measurement.
4. **Conformity Assessment** – verifying that products and services meet specified requirements.
5. **Enforcement** – monitoring markets and addressing non-compliance to protect consumers and businesses.

Within the UK Quality Infrastructure, metrology provides the measurement capability that underpins activities such as standardisation, accreditation, conformity assessment, and regulation. However, metrology is itself a broader discipline than the metrological functions captured within the UKQI, encompassing a wider range of scientific, industrial, and research activities.

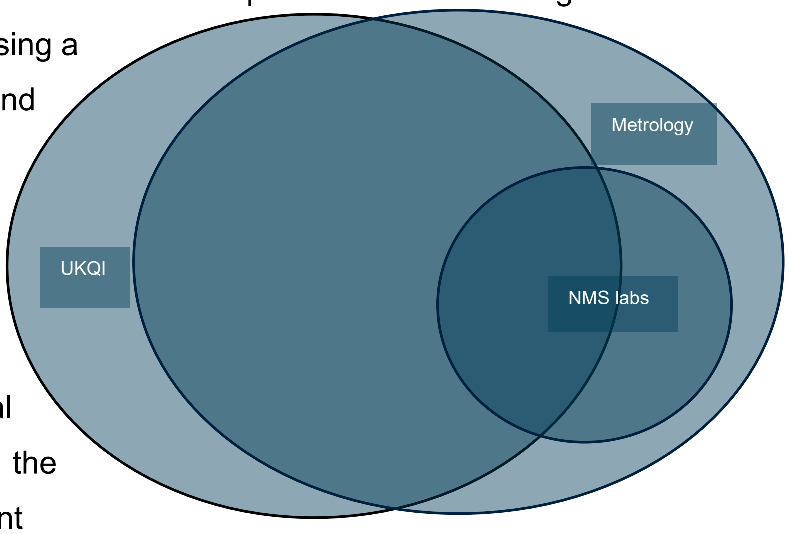


Figure 1 Position of the NMS Laboratories

At the centre of the UK's metrology landscape sits the National Measurement System (NMS) and its labs, which represent the UK's formal measurement infrastructure and the source of traceable measurement standards. The NMS operates as a distributed network designed to ensure the reliability, traceability, and international comparability of measurements across the UK. In doing so, it provides a critical foundation for the metrological functions of the UKQI while also supporting measurement activities across the wider economy and research base.

Beyond the NMS, metrological activities are undertaken by private-sector calibration laboratories, industrial measurement functions, commercial testing services, and research organisations. Equally, the UKQI encompasses functions that extend beyond metrology, including standardisation, accreditation, conformity assessment, and enforcement. As shown in Figure 1, the NMS should therefore be viewed as the UK's formal measurement infrastructure, occupying a central position within the wider metrology landscape and underpinning the measurement-related functions of the UK Quality Infrastructure.

As a result, the economic benefits generated by the NMS can extend beyond metrology itself and ultimately propagate throughout the wider quality infrastructure and the broader economy. At its core are the top-level laboratories funded by the Department for Science, Innovation, and Technology (DSIT), including the National Physical Laboratory. These labs maintain primary standards and support a wider community of accredited laboratories that provide traceable calibrations and certified reference materials. This structure enables comparability of measurements across diverse organisations, even those not directly using top-level labs. By linking UK measurements to the International System of Units (SI), the NMS participates in international key-comparison exercises to uphold Mutual Recognition Arrangements (MRAs), which remove technical barriers to trade and support global market access.

This system-wide role of measurement is reflected in a substantial economic footprint, as evidenced by multiple studies quantifying its contribution to UK output and activity. Measurement is a foundational economic infrastructure with significant economy-wide impact. Evidence over the past 25 years shows that it contributes around 0.8% of UK GDP (≈£24 billion in 2025 terms) and underpins large volumes of economic activity, with over £622 billion of trade relying on measured quantities. Measurement activity is also deeply embedded across the economy: around 6.3% of UK employment involves measurement-related roles, contributing roughly £85 billion of output. Direct users of the National Measurement System (NMS) spend approximately 5% of their turnover on measurement annually, equating to £7.7 billion of spending each year. Followed by impacts that cascade through nearly 75,500 organisations via calibration supply chains. This highlights measurement as a critical enabling infrastructure for productivity, trade, and reliable economic decision-making. [3]

1.2 Conceptual Foundations

Why a Macroeconomic approach?

Primary measurement standards function as a single source of durable public information (a single source of truth) and thereby serve as an anchor for the credibility of the torrent of fleeting bits of measurement information generated each day by a vast array of different businesses and organisations. Therefore, measurement can be viewed as an "**infra-technology**". This is supported by a well-established body of literature that identifies measurement as a foundational "infra-technology" with economy-wide effects. For instance, BEIS (2017), building on earlier contributions by Tassey (1982) and Estibals (2012), emphasises that measurement underpins productivity, innovation, and regulatory compliance across almost all sectors of the economy and is important for both product and process innovation. This broader, system-level perspective is also reflected in the work of Blind, who highlights the role of standards and measurement in shaping firm behaviour, regulation, and innovation across both micro- and macroeconomic contexts. Moreover, standards and measurement frameworks embed reliable information into production and exchange systems by codifying best practice and ensuring consistency in how quantities are defined and assessed. In this sense, **good measurement is fundamentally** about **improving the quality of information available** to economic actors. There is clear precedent for analysing these effects at the macroeconomic level. For example, the DTI (2005) study on the economics of standards shows how the accumulation of standards contributes to productivity growth across the economy. As production systems become increasingly automated, particularly with the rise of Industry 4.0 and AI-driven processes, the importance of high-quality, standardised measurement information becomes even greater, as decisions are increasingly made by systems rather than individuals. Additionally, the benefits of advances in measurement science are widely shared across the economy and are not fully captured by any single firm or sector. This leads to systematic underinvestment and causes market failure, providing a strong economic rationale for why

[3] PA review (Review of the Rationale for and Economic Benefit of the UK National Measurement System) is one of the oldest studies that calculate and analyse the economic benefit of measurement science. It was published in 1999 and it is also the last time the NMS saw a large-scale review.

most governments will invest in measurement science. Hence, this modelling approach does not represent the first attempt to adopt a macroeconomic perspective on measurement infrastructure.

Measurement, Information, and Operational Decisions

Measurement can be understood as the process of generating information that reduces uncertainty. Drawing on Shannon (1948), uncertainty can be framed in terms of entropy, where more precise measurement reduces the range of possible outcomes and thus the "surprisal" associated with decisions. In this context, advances in metrology increase the informational content of measurements, enabling more accurate operational decisions. This informational role is not just conceptual but economically significant in the sense that by reducing uncertainty, better measurement lowers the likelihood of Type I and Type II errors, reduces waste, and therefore, enables tighter process control. Qureshi and King (2025) suggest that measurement generates substantial returns: for example, improved measurement capability can yield around £1.30 in benefits for every £1 spent, reflecting its role in improving decision accuracy and efficiency. At the economy-wide level, this translates into a large footprint, with billions of pounds of output, trade, and employment relying on measured information. [4]

In practice, this reduction in uncertainty plays out at the firm level through day-to-day decision-making. Firms continuously make operational decisions—such as whether output meets specifications, whether a process is functioning correctly, or whether equipment requires maintenance—based on measured data. The reliability of these decisions depends directly on the quality of the underlying measurement information. When measurement uncertainty is high, decision-making becomes less dependable, increasing the likelihood of errors. Conversely, more precise and reliable measurements reduce uncertainty and improve the accuracy of day-to-day decisions. This relationship provides the foundation for modelling conformance testing, where imperfect measurement leads to decision errors in the form of false rejections and false acceptances of output.

Conformance testing and the engineers who perform it play an important role in reducing operational inefficiencies by generating information about the state of production processes. This function is closely related to the economic concept of x-inefficiency, introduced by

[4] Found in 'A framework for assessing the economic significance of NPL's spillovers'

Leibenstein (1966), which explains why firms often operate above their minimum possible costs because of organisational slack, weak incentives, and imperfect monitoring. By revealing production defects and equipment malfunctions that might otherwise go unnoticed, conformance testing helps firms reduce these inefficiencies and operate closer to the efficient frontier. A complementary perspective is provided by Sims' (2003) theory of rational inattention, which argues that firms face limits in their ability to process information and therefore allocate their attention selectively. As a result, production errors, equipment failures, and quality deviations may go unnoticed, allowing inefficiencies to persist. In both frameworks, conformance testing helps overcome informational constraints by generating reliable signals about the state of production and directing attention towards economically relevant problems.

Conformance testing acts as a powerful shield against these internal inefficiencies whereby CT Engineers serve as an objective quality control check by ensuring that every product meets exact, predetermined standards. This process uncovers hidden mistakes and prevents workers from doing subpar work, pushing the business to maximise efficiency. Moreover, the optimal testing model can be viewed as balancing the benefits of acquiring additional information against the costs of obtaining it, with the optimal frequency of testing emerging where the marginal value of information equals its marginal cost.

Conformance testing also serves as a mechanism that mitigates technological uncertainty. Following Arrow (1962) and MacMinn and Holtmann (1983), technological uncertainty is conceptualized as imperfect information regarding production processes or output quality. This lack of clarity dampens long-term investment (Ye, Paulson, and Khanna, 2024), degrades feedback mechanisms, and increases decision errors—ultimately generating sub-optimal productivity. Conformance testing counters these challenges by enhancing measurement precision and reliability, thereby strengthening the informational feedback loop from the production floor. By allowing firms to accurately distinguish compliant from defective output, testing reduces decision errors and diminishes structural waste. Consequently, firms can operate closer to the efficient frontier, proving that conformance testing functions not only as a quality control mechanism, but also as a direct driver of productive efficiency.[5]

Beyond internal decision-making, the quality of measurement information also scales up to affect trade and market access. Reliable and internationally recognised measurement reduces transaction costs in trade. Confidence in measurement equivalence allows goods to be “measured once and accepted everywhere”, avoiding costly duplication of testing and conformity assessment across borders. This is supported by institutional mechanisms such as Mutual Recognition Arrangements (MRAs) and international key comparison exercises, which ensure the comparability and traceability of measurements across national systems. By lowering informational frictions, these frameworks facilitate trade and the efficient operation of global value chains.

Modelling Decision Errors in Production

This relationship between measurement quality and decision-making can be formalised through a framework of conformance testing, where high-quality measurement reduces uncertainty in classification decisions, a link formalised in the microeconomic model developed by King and Nayak (2023). Their work is built on the notion that the conditional probabilities of Type I (false positive) and Type II (false negative) errors are decision errors shaped by measurement quality. High-quality, traceable calibrations reduce measurement uncertainty, allowing decision makers to re-optimise thresholds and lower total loss from misclassification.

Building on this, King and Nayak (2025) take a theory-based approach to incorporate the national quality infrastructure into a classic macroeconomic model, namely, the Solow Model. Solow’s work in the 1950s explained that an economy can raise output per person by adding more capital per worker (capital deepening), but only up to the steady state value of capital intensity. The Solow model assumes full employment, so the size of the labour force rises with population. A key equation within the Solow model is one which describes the change in capital stock through new investment financed by saving and depreciation

[5] In contrast to operational risks, there also exists market uncertainty which arises from unpredictable shifts in the external economic environment. It is characterized by imperfect information regarding consumer demand functions, volatile factor input prices, and the strategic behaviour of competitors (Dixit and Pindyck, 1994). While technological uncertainty disrupts the internal production process, market uncertainty distorts price signals and revenue forecasting. Consequently, even a firm operating at maximum technical efficiency faces the risk of a market mismatch, where final outputs cannot be cleared at profitable prices.

that reduces the existing capital stock. Solow showed that capital accumulates until the economy's capital intensity reaches a steady state. At this point investment equals the effects of depreciation. Solow showed that there exists a specific savings rate that maximises steady-state consumption per worker known as the **Golden Rule savings rate**, balancing the trade-off between current and future consumption. Saving less leaves the economy under-invested, reducing output, saving more forces excessive resources into investment at the expense of consumption.

This concept provides a useful analogue for understanding the macroeconomic role of conformance testing within the measurement infrastructure. Just as the Golden Rule savings rate identifies the equilibrium which maximises consumption per worker, the determination of the optimal testing frequency follows the logic of the "Golden Rule," where the equilibrium level of testing is defined by the condition that marginal benefits equal marginal costs, maximising consumption per worker. This level is associated with raised productivity due to greater share of functioning capital and reduced defective output.

This equilibrium is not fixed and depends on underlying parameters such as error probabilities and effectiveness of testing. This is where comparative statics becomes relevant. Changes in parameters such as an increase in the probability of Type I errors reduce the marginal benefit of testing and shifts the equilibrium towards a lower optimal testing frequency. In this way, comparative statics provides a useful tool for analysing how changes in measurement quality translate into changes in economic outcomes.

This model can also be interpreted through a demand and supply lens. The demand for testing is derived from the marginal benefits of improved reliability, while the supply reflects the marginal cost of performing tests. Improvements in measurement quality, such as those enabled by high-quality, traceable calibrations, reduce the probability of Type I errors (p), thereby increasing the effectiveness of testing. In graphical terms, this shifts the demand curve for testing upwards, as each additional round of testing yields greater net benefit. Conversely, higher measurement uncertainty (higher p) shifts the demand curve downwards, lowering the equilibrium level of testing.

In this manner the framework allows the model to be used for **comparative statics**, where changes in parameters generate shifts in the equilibrium level of testing, reliability, and output. It is precisely this feature that enables the construction of **counterfactual scenarios** like analysing how reduced access to traceable calibrations (in the absence of a domestic NMI) would alter testing activity and economic performance across the system. Hence, the Optimal Testing Model adds to the measurement infrastructure framework by quantifying the changes in demand for conformance testing attributable to public funding of NMS laboratories and their provision of high-quality calibrations.

1.3 Purpose and Scope of the Report

This report aims to unpack and provide an intuitive explanation of the macroeconomic model presented in *King & Nayak (2025)* and discuss the calculator developed for the model. The work on the macroeconomic model is built on two earlier reports:

- Fennelly, C. (2021). *Quantifying Measurement Activity in the UK*. NPL Report. IEA7
- King, M. & Nayak, S. (2023). *An Economic Model for the Value Attributable to High-Quality Calibrations by Reducing Mistakes in Conformance Testing*. NPL Report. IEA19

The framework presented in *King & Nayak (2025)* comprises of four principal components and adding a final component gives a holistic representation for the framework of the 'traceability benefits' channel of impact:

1. **Gravity Model**
   The model is termed a gravity model because, like gravity in physics, interactions increase with economic size and decrease with distance. It therefore includes GDP, population, and the straight-line distance between London and a country's largest city. The model estimates how demand for precise calibrations falls with distance from NMS laboratories and uses this relationship to analyse scenarios in which domestic NMS capability is reduced or absent. In such scenarios, firms reduce their use of precise calibrations and rely more heavily on basic calibration services.
2. **Calibration and Fanout Model** (work-in-progress)
   The Calibration and Fanout Model characterise how traceable measurement information is generated, distributed, and sustained across the calibration chain in equilibrium. It explains why calibration activity expands as it moves from top-level

institutions (such as NPL) to lower tiers, resulting in an exponential increase in the volume of calibrated instruments. The model also examines how operational factors—such as turnaround times and pricing—propagate through the chain, shaping the availability, quality, and economic value of calibrations for end users.

3. **Basic and Precise Calibrations Model**

   It explains that measurement errors consist of two components: Random error (that can be averaged out); and Systematic error (that can only be reduced via precise calibration). The model distinguishes two calibration types:

   - Precise calibrations (from NMIs): These minimise/ remove systematic error and significantly improve accuracy.
   - Basic calibrations: These leave residual systematic error (linked to the Test Accuracy Ratio, which corresponds to extra uncertainty introduced through less accurate calibrations which could be as much as 25% of the baseline uncertainty). [6]

4. **Type I and Type II Errors Model**

   It describes how measurement uncertainty in conformance testing leads to decision errors, specifically false positives (Type I errors) where conforming output is incorrectly rejected, and false negatives (Type II errors) where defective output is incorrectly accepted. The model formalises the trade-off between these errors and shows how their rates depend on the Relative Standard Deviation (RSD) of measurements and the chosen inspection regime, with direct implications for waste, reliability, and economic efficiency. [7]

5. **Optimal Level of Testing Model**

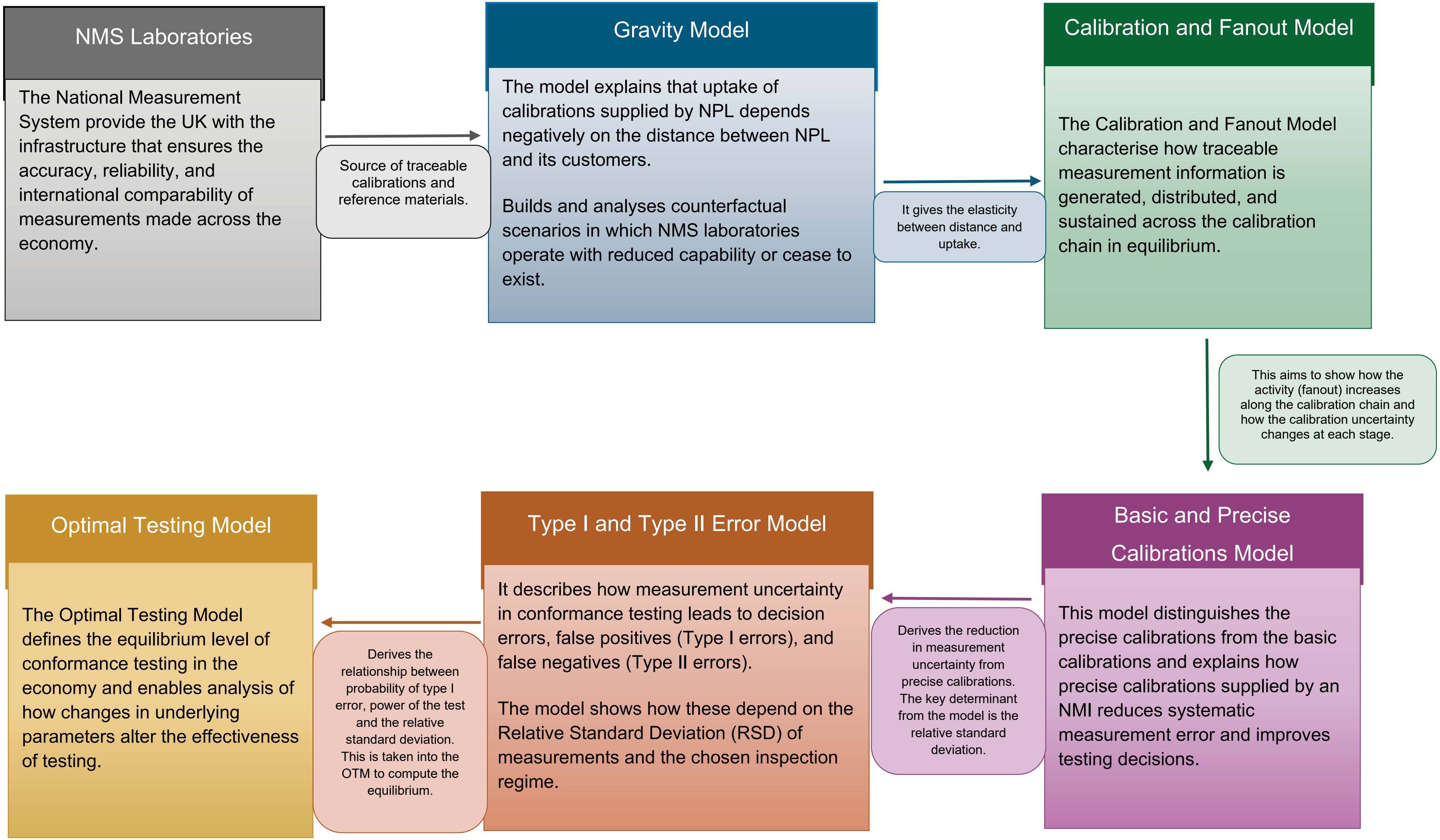
   The Optimal Testing Model defines the equilibrium level of conformance testing in the economy and enables analysis of how changes in underlying parameters alter the effectiveness of testing. Crucially, it provides a platform for quantifying the changes in demand for conformance testing attributable to public funding of NMS laboratories and their provision of high-quality calibrations. Through its integrated calculator, the model illustrates how variations in underlying parameters affect both

[6] TAR is the comparison between the accuracy of a tool (Unit Under Test or UUT) and the reference standard used to calibrate it. Metrology labs aim for a minimum TAR of 4:1. This implies that the extra uncertainty introduced through less accurate calibrations could be as much as 25% of the baseline uncertainty.

[7] It is the deviation that occurs under the hypothesis that parts of a machinery are defective. It depends on the expected uncertainty of measurement process.

the supply and demand of measurement testing, offering valuable insight into the broader economic impacts of calibration quality and testing infrastructure.

These system of models embody the channel of impact for the NMS, ‘Delivering the Measurement Infrastructure’. The figure below shows the link between each of these models and the Annex B explains each of these models and its key features in detail.


NMS Laboratories
The National Measurement System provide the UK with the infrastructure that ensures the accuracy, reliability, and international comparability of measurements made across the economy.
Source of traceable calibrations and reference materials.
Gravity Model
The model explains that uptake of calibrations supplied by NPL depends negatively on the distance between NPL and its customers.
Builds and analyses counterfactual scenarios in which NMS laboratories operate with reduced capability or cease to exist.
It gives the elasticity between distance and uptake.
Calibration and Fanout Model
The Calibration and Fanout Model characterise how traceable measurement information is generated, distributed, and sustained across the calibration chain in equilibrium.
This aims to show how the activity (fanout) increases along the calibration chain and how the calibration uncertainty changes at each stage.
Basic and Precise Calibrations Model
This model distinguishes the precise calibrations from the basic calibrations and explains how precise calibrations supplied by an NMI reduces systematic measurement error and improves testing decisions.
Derives the reduction in measurement uncertainty from precise calibrations. The key determinant from the model is the relative standard deviation.
Type I and Type II Error Model
It describes how measurement uncertainty in conformance testing leads to decision errors, false positives (Type I errors), and false negatives (Type II errors).
The model shows how these depend on the Relative Standard Deviation (RSD) of measurements and the chosen inspection regime.
Derives the relationship between probability of type I error, power of the test and the relative standard deviation. This is taken into the OTM to compute the equilibrium.
Optimal Testing Model
The Optimal Testing Model defines the equilibrium level of conformance testing in the economy and enables analysis of how changes in underlying parameters alter the effectiveness of testing.

The framework links the constituent models through a clear and structured flow of information and dependencies. At its foundation, **NMS laboratories provide calibrations that are traceable to national standards**, acting as the primary source of traceability, calibration, and reference materials. To model the impact generated by these laboratories, the framework begins with the '**Gravity model**'. This model represents the first tier of the calibration chain and explains how firms' uptake of calibrations depends negatively on distance from NPL. It **estimates the elasticity between distance and uptake**, while also enabling analysis of counterfactual scenarios in which NMS laboratory capability is reduced or absent. The next part of the framework characterises how traceable measurement information is generated, distributed, and sustained across the calibration chain in equilibrium through the **Calibration and fanout model**. It captures the structure of the 'fanout' of calibrations throughout the supply chain and determines how initial calibration activity propagates across different tiers of economic activity. The next model '**Basic and Precise Calibrations Model**' distinguishes between basic and precise calibrations by quantifying how precise calibrations reduce measurement error, expressed through reductions in relative standard deviation. The reduction in measurement uncertainty is subsequently input into the '**Type I and Type II Error Model**'. This model links measurement precision to decision outcomes in conformity testing, describing how uncertainty increases false positives (Type I errors) and false negatives (Type II errors), and how these depend on the relative standard deviation of measurements and the inspection regime. Finally, these relationships are passed to the '**Optimal Testing Model**', which uses them to determine the equilibrium level of testing activity. The model derives how the probabilities of decision errors and the standard deviation of measurements jointly influence testing behaviour and the key determinants that shape optimal measurement activity.

Having given the context of the framework, this report focuses solely on the fifth component of the system to make the underlying logic of the model more accessible and to demonstrate its practical application via the calculator. Subsequent studies will unpack the other models and bring them together.

The theoretical framework for determining the optimal level of testing in the economy is developed by adapting the Solow growth model and using the equilibrium condition where marginal benefits equal marginal costs. The concept of the "Golden Rule" from the Solow–Swan model is applied to derive an analogous rule for optimal testing. The model, therefore, aims to show how conformance testing is essential for an economy to function efficiently. The report also introduces a practical calculator to illustrate how various components of the model are derived and also compute the comparative statics which shows the implications of changing a basic parameter such as the probability of type I error.

The structure of this report is as follows: Section 2 sets out the model, introducing production under imperfect reliability and the role of conformance testing, along with the testing functions and the equilibrium components of the Solow-based framework. Section 3 develops the marginal cost and marginal benefit of additional testing. Section 4 defines all the parameters, variables and derived quantities used in the model and explains the interrelationships. Section 5 presents the solution for the optimal level of testing and provides a numerical illustration of the model, followed by comparative statics using the calculator tool. This examines the counterfactual where NMS labs no longer exist and the effect on the components of the model. Section 6 provides the conclusion, summarising the analysis and discussing the implications of the model's derived components.

2 Setup of the Model

2.1 Production in the economy

This subsection sets out the production framework, beginning with the representation of economic output under imperfect reliability of production and introducing the role of conformance testing in maintaining reliability of production processes.

Time t is continuous and measured in years, and capital is measured in millions. There is full employment, and the size of the workforce is set by the population. Economic output, Y, depends on combining labour, L , with capital equipment, K, according to the economy's production function: $Y = \mathrm{F}(L, K)$. Production is represented by a Cobb-Douglas production function with constant returns to scale. If production processes were perfectly reliable and everyone was employed in production, then the economy's output is given by:

$$Y = AL^{\alpha}K^{\beta}$$

where, A and α, β are fixed parameters of the model: A is the economy's Total Factor Productivity (TFP); α is the elasticity of output with respect to labour; β is the elasticity of output with respect to capital; and constant returns to scale implies $\alpha + \beta = 1$.

The malfunctioning in capital effect the production process by effecting both the capital and labour and can be shown as

$$Y = A(vL)^{\alpha}(vK)^{\beta} = Av^{\alpha+\beta}L^{\alpha}K^{\beta} = AvL^{\alpha}K^{\beta}$$

where v is denoted as the reliability of production. Testing output to check the reliability of production means there will be a proportion of output that passes through the testing and a small proportion that incorrectly gets thrown away as defective output (regret rate) also known as Type I error. This can be shown as:

$$Y = A(1-\theta)vL^{\alpha}K^{\beta}$$

where effective TFP (B) is equal to $A(1-\theta)v$. Note that in the model we cannot observe the total TFP (A) but we are able to observe the effective TFP(B)

The intensive form of the production function can be written as

$$y = A(1-\theta)vk^{\beta} \tag{1}$$

where y is the output per worker (labour productivity), and k is the capital intensity (capital per labour).

Throughout the document, d and Δ are used interchangeably for the derivative, and defected output may be referred to as ‘bad’.

2.2 Testing Functions

2.2.1 Production for the input and output of testing services

Given capital can malfunction and produce defective output, it is subject to conformance testing by CT engineers. In this framework, it is the output—not the machine itself—that is directly tested. Say, a machine produces widgets according to a specified standard. Its functionality cannot be assessed in isolation; it is only through inspecting the widgets that it can be determined whether the capital is operating correctly. If the output meets the required characteristics, the machine is deemed to be functioning properly; if it does not, the capital is considered malfunctioning. In such cases, engineers rely on the evidence from output testing to trace the issue back to its source, reset or adjust the relevant component, and restore the system so that it continues to produce conforming output.

The setup computes the marginal benefit and cost to see the impact of an additional round of testing. The model assumes that there is some existing level of testing in the economy. Each engineer is assigned a portion of capital say a specific production plant and is responsible for monitoring its output for any signs of malfunctioning. Upon detecting a fault, the engineer intervenes to reset the machinery and equipment associated with the issue. In essence, this model can be thought of as **process control informed by conformance testing.**

The model set out below looks at how much testing is feasible by considering the number of rounds required to ensure that all the capital is functioning. In other words, how many CT engineers can be hired and the amount each can test, determines the quantity of testing.

Production function for the testing services is written as

$$\underbrace{(n \times k)}_{demand\ for\ testing} = \underbrace{(a \times e)}_{supply\ for\ testing} \tag{2}$$

where:

- n denotes the frequency of testing (rounds per year), which with a fixed capital intensity (steady state) implies a proportional increase in the fraction of the workforce employed in conformance testing and process control.
- k is the amount of capital per labour (capital intensity).

- a is the amount of capital a CT engineer ensures is functioning correctly in a year, and $1/a$ is the time it would take to ensure £1m worth of capital is functioning correctly. It is also known as the pace of the CT engineer.
- e represents the CT engineers as a proportion of labour force that engages in testing. CT engineers are mobile and operate in a global labour market. The wages paid to them are from the share of output produced by the productive labour force. Their wages set at a global level and so treated as exogenous. Their remuneration is paid out of the output produced by the domestic productive workforce. The model assumes a closed economy with respect to all other workers: there is no economic migration, and labour supply in the domestic sector is fixed. The only exception is CT engineers, who can move freely across borders and whose wages are unaffected by domestic UK labour market conditions.
- The LHS denotes the demand for testing i.e. how much capital needs to be tested and the RHS denotes the supply of testing i.e. how many engineers are available to test the capital and how quickly can it be tested.

2.2.2 Tax and Spend equation

As described above CT engineers belong to the non-output producing sector so their wages are paid through the share of output produced by output producing labour force. From here a concept called 'reliability tax' is derived. This is not a tax in the traditional sense but like an overhead to fund the wages of the CT engineers. An example here is, Human Resources (HR) plays a vital role in managing an organisation, even though it doesn't directly contribute to the production of output. The salaries of HR staff are typically covered through the organisation's overhead costs.

The 'reliability tax' (τ) is considered to be proportional to the capital. The more capital there is, the more testing is required, and hence greater share of output goes to funding the wages of CT engineers. Hence, 'reliability tax' is applied to the capital intensity, and the tax equation can be written as $\tau \times k$.

The wages paid to the CT engineers can be denoted as the spend equation and written as $e \times w$. The tax equation should equate to the wages paid to the CT engineers as it ensures that the wages of the workers correspond to the taxation funds generated.

The Tax and spend equation can, therefore, be written as:

$$\underbrace{w \times e}_{spend\ side} = \underbrace{\tau \times k}_{tax\ side} \tag{3}$$

The equation above denotes that the total spending on conformance-testing engineers must equal the revenues raised through the reliability tax.

2.2.3 Testing intensity equation

The testing intensity (m) can be denoted as the income to the CT engineers. It can be written as

$$m = \frac{e \times w}{y}$$

It is simply the proportion of output that goes as the total wages to the CT engineers.

2.2.4 Relationship of 'reliability tax' to frequency of testing

How does the 'reliability tax' (τ) impact the frequency of testing? To find a relationship between τ and n let's bring the expression of the production function for the supply of testing and the tax and spend equation together.

$$n \times k = a \times e$$

$$e \times w = \tau \times k$$

Substituting the value of k in the production function. The expression can be written as

$$a \times \tau = w \times n$$

$$\frac{a}{n} = \frac{w}{\tau}$$

Where $^{a}/_{n}$ is the portion size of capital that goes to each CT engineer in a round of testing[8]. Rearranging the expression as:

$$\tau = \frac{w}{a} \times n \tag{4}$$

[8] $n \times k = a \times e \rightarrow {}^{k}/_{e} = {}^{a}/_{n}$ → Equivalent to the portion size of capital

It shows that 'reliability tax' (τ) is proportional to the frequency of testing given wages and portion size remain constant.

It can also be expressed as $\tau = \frac{w}{\rho}$, reinstating a negative relationship between 'reliability tax' and portion size of capital going to a CT engineer.

2.2.5 Regret and Detection rate

Regret rate (θ) also known as the type I error is the good output thrown away accidently. How would the regret rate be impacted by the frequency of testing?

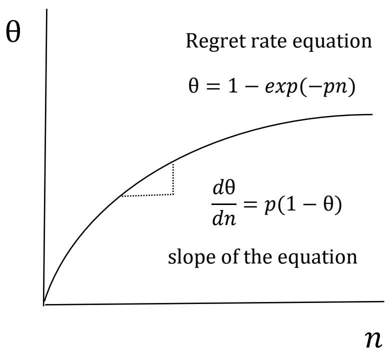


Figure 2 Relationship of frequency of testing and Regret Rate

The figure shows the relationship between frequency of testing and the regret rate. It shows that as the frequency of testing increases, a fraction of good output falls prey to type I errors.

The differential equation defining regret rate is simply the slope equation and written as:

$$\frac{d\theta}{dn} = p(1 - \theta) \tag{5}$$

where p is the probability of making a type I error, θ is the regret rate (the rate at which output is incorrectly scrapped). Integrating the expression gives an expression for the regret rate which can be expressed as, $\theta = 1 - exp(-pn)$

Using a finite-difference approximation, where $d\theta$ and dn are replaced with small discrete changes, the expression is written as:

$$\Delta\theta = p \times (1 - \theta) \times \Delta n \tag{6}$$

This expression is a key component of the model. It captures the idea that, although a share $(1 - \theta)$ of output passed testing as compliant in the previous round, there remains a probability *p*p of a Type I error in the subsequent round. That is, with every additional round of testing, a proportion of good output falls prey to type I errors.

The detection rate (ϕ) is the rate at which defective output is detected, and the malfunctioning part of capital is reset. It has a similar differential expression to the regret rate.

$$\frac{d\phi}{d\mathrm{n}} = q(1-\phi) \tag{7}$$

where q is the probability of correct detection (power of the test). Similar to above using finite-difference approximation, the expression can be written as:

$$\Delta\phi = q \times (1-\phi) \times \Delta n \tag{8}$$

$(1-\phi)$ is what previously couldn't be detected as bad and so could not be reset. With each round of testing, a fraction of bad undetected output is correctly detected and malfunctioning capital producing the bad output is reset. From the differential, the expression for the detection rate is simply $\phi = 1 - \exp(-qn)$.

2.3 Solow model & Equilibrium Components

2.3.1 Equilibrium level of capital intensity

This is also known as the Markov model for capital intensity (k^*). The equilibrium level of capital intensity is derived from the steady state of the Solow model. The steady state happens when the amount of capital per worker remains constant over time (inflow is equal to outflow of capital). That is, the new investment being added to the economy is just enough to cover depreciation and dilution from population growth.

In our model setup, the existence of the reliability of tax on capital would imply that net output per worker is:

$$y^* - (\tau \times k^*)$$

In the Solow model, the savings rate (s) is assumed to be constant. And gross savings are denoted by sy. Consequently, gross savings in our model can be written as:

$$s(y^* - \tau \times k^*)$$

The expression above is easily comprehended when compared to take home salary where all taxes are deducted from the gross salary and a proportion of the salary left is saved.

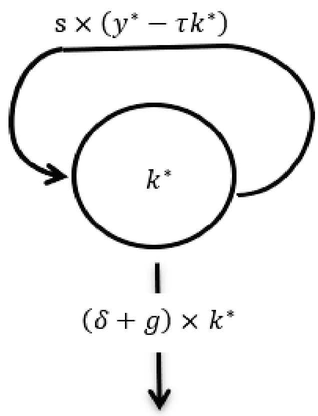


Figure 3 Capital intensity in the steady state

In the steady state, since capital per worker (k^*) remains constant over time, i.e., at this point investment equals depreciation and dilution (inflow is equal to outflow of capital). This can be explained through the diagram that represents a type of a Markov process where the inflow is the savings that get added back to the capital as investment and the outflow is the depleted capital and the reduction in capital as a result of increased growth in workforce.

The inflow can be written as $s = (y^* - k^* \times \tau)$ and the outflow can be written as $(g + \delta)k^*$ where depreciation of capital is denoted by δ and growth of population by g.

In the steady state, since inflow and outflow of capital are equal, the expression, therefore, can be written as:

$$s(y^* - \tau \times k^*) = (g + \delta)k^*$$

The above expression confirms that if there were no CT engineers the maximum saved would be simply sy (LHS).

Rearranging, we get the following expression for the steady state capital intensity:

$$sy^* = (g + \delta + s\tau)k^*$$

$$\boxed{k^* = \frac{sy^*}{(g + \delta + s\tau)}} \qquad (9)$$

It tells us the amount of capital per worker that can be sustained in the long run once all leakages from the capital stock are accounted for.

2.3.2 Equilibrium level of testing intensity

Substituting the expression for the equilibrium capital intensity in the tax and spend equation:

$$e \times w = k \times \tau$$

$$e \times w = \frac{s\tau}{(g + \delta + s\tau)} \times y^* \qquad (10)$$

Rearranging the expression, gives:

$$\frac{e \times w}{y^*} = \frac{s\tau}{(g + \delta + s\tau)}$$

And since $m = \frac{e \times w}{y}$, the testing intensity can be written as:

$$\boxed{m^* = \frac{s\tau}{(g + \delta + s\tau)}} \qquad (11)$$

Equation (11) shows that the equilibrium testing intensity m^*, is a function of the savings rate, s. Treating s as exogenous is a "fudge" borrowed from Solow because it assumes consumers never alter their saving behaviour when testing conditions change. Despite this limitation, the model maintains this simplification to keep the mathematical framework tractable.

2.4 Markov Model for Reliability of Production

The diagram below represents a kind of Markov model for the reliability of production. It shows the flows inside and outside of the functioning and malfunctioning capital. It shows two buckets: one with functioning capital and the other malfunctioning capital and it explains the inflows (positive contributions) and outflows (leakages) from these buckets, when the system is in equilibrium. Initially all capital is functioning and falls into the first bucket.

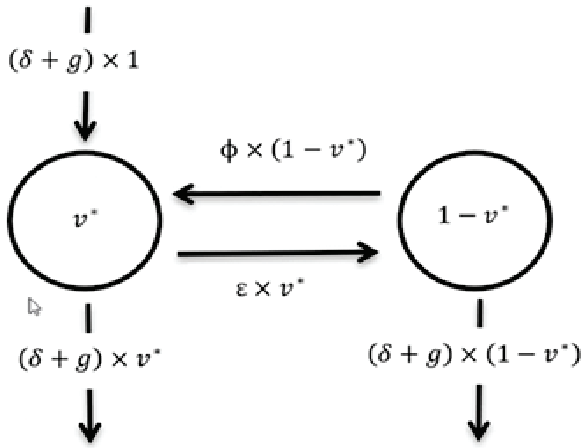


Figure 4 Flows of probability in and out of v^* and $1 - v^*$

There are inflows coming from the gross investment and then there are two types of outflows: one is in the form of depreciated capital and capital reductions due to growth in workforce; and the second one is functioning capital that flips and produces defected output. Functioning capital can naturally flip to producing defected output due to entropy in production processes[9]. From the malfunctioning capital, the inflow is simply the functioning capital that flips, and the two outflows are: depreciation; and the malfunctioning capital whose output gets detected and part of the capital malfunctioning gets reset, so flows back into the functioning capital bucket. In equilibrium these inflows and outflows from each bucket must be equal. Using this condition, we can get an expression for the equilibrium level of reliability for capital.

$$\varepsilon \times v^* = \phi \times (1 - v^*) + (\delta + g) \times (1 - v^*)$$

Solving for v

$$\varepsilon v^* = \phi - \phi v^* + \delta - \delta v^* + g - g v^*$$

$$\varepsilon v^* + \phi v^* + \delta v^* + g v^* = \phi + \delta + g$$

[9] All systems naturally wear down, drift, degrade, or move toward disorder unless energy is put in to maintaining them.

$$\boxed{v^* = \frac{\delta + g + \phi}{\delta + g + \phi + \varepsilon} \quad where\ v < 1} \tag{12}$$

The derivative of v^* w.r.t ϕ can be written as:

$$\frac{dv^*}{d\phi} = \frac{1 - v^*}{\delta + g + \phi + \varepsilon}$$

For simplicity in taking a derivative, consider the following

$$v = \frac{N}{D} \rightarrow N = vD$$

where $N = \delta + g + \phi$ and $D = \delta + g + \phi + \varepsilon$

Taking the derivative of v w.r.t ϕ

$$\frac{dv}{d\phi} = \frac{D - N}{D^2} = \frac{D(1 - v)}{D^2} = \frac{1 - v}{D}$$

$$\frac{dv}{d\phi} = \frac{1 - v}{\delta + g + \phi + \varepsilon}$$

How is the reliability of production affected by additional testing?

Differential approach can be used to express the change in reliability as a result of a change in the frequency of testing. The reliability v^* depends on n, so its rate of change is given by the partial derivative $\frac{\partial v^*}{\partial n}$. For a small increment in n (an additional round of testing), the corresponding change in v can be approximated by:

$$\Delta v^* \approx \frac{\partial v^*}{\partial n} \cdot \Delta n$$

Multiplying and dividing by v^*

$$\frac{dv^*}{v} = \frac{1}{v^*}\frac{\partial v^*}{\partial n} \times dn$$

Multiplying and dividing by n to get the elasticity $(\mathcal{E}(v, n))$ of v w.r.t n.

$$\frac{dv^*}{v^*} = \underbrace{\frac{n}{v^*}\frac{dv^*}{dn}}_{\mathcal{E}(v,n)} \times \frac{dn}{n} \tag{13}$$

Chain rule of differentiation on ${dv^*}/_{dn}$ can be used to derive an expression in terms of ϕ as it is difficult to obtain the expression for the partial derivative of v^* w.r.t n directly.

$$\frac{dv^*}{d\mathrm{n}} = \frac{dv^*}{d\phi} \times \frac{d\phi}{d\mathrm{n}}$$

Since we know both the products, the expression can be written as:

$$\frac{dv^*}{d\mathrm{n}} = \frac{1 - v^*}{\delta + g + \phi + \varepsilon} \times q(1 - \phi)$$

The expression for the change in reliability w.r.t to frequency of testing can be written as:

$$\frac{dv^*}{d\mathrm{n}} = \frac{q(1 - \phi)(1 - v^*)}{\delta + g + \phi + \varepsilon} \tag{14}$$

Substituting this back into equation 9, the expression can be written as:

$$\frac{dv^*}{v^*} = \frac{n}{v^*}\frac{q(1 - \phi)(1 - v^*)}{\delta + g + \phi + \varepsilon} \times \frac{dn}{n} \tag{15}$$

This expression shows how sensitive the reliability of production v^* is to a small increase in testing frequency n. The left-hand side $\frac{dv^*}{v^*}$ is the percentage increase in reliability and the right-hand side is the product of proportional change in testing, and elasticity of reliability w.r.t testing. Imagine you service your car more often, $\frac{dn}{n}$ is how much more frequently you service it, the middle fraction is how effective servicing is at preventing breakdowns, and $\frac{dv^*}{v^*}$ is how much less likely your car is to break down. So, the equation tells you, *"A 10% increase in servicing improves reliability by (elasticity × 10%)"*.

This equation will be used later when we derive the marginal benefits equation for an additional round of testing.

3 Costs and Benefits of Additional Testing

3.1 Average and Marginal Cost of Testing at the Optimum

Let's start with the production function of testing.

$$a \times e = n \times k$$

Differentiating both sides w.r.t n.

$$a \frac{de}{dn} = k \frac{dn}{dn} + n \frac{dk}{dn}$$

Pace of testing is not affected by n so comes out as a constant. Product rule of differentiation applies on the term kn.

$$a \frac{de}{dn} = k + n \frac{dk}{dn}$$

At the steady state $\frac{dk}{dn} = 0$, therefore:

$$a \frac{de}{dn} = k$$

Multiplying and diving by n

$$a \frac{de}{dn} = k \times \frac{n}{n} \;\rightarrow\; a \frac{de}{dn} = \frac{a \times e}{n} \;\rightarrow \frac{de}{dn} = \frac{e}{n}$$

Multiplying and dividing by ω

$$\omega \times \frac{de}{dn} = \frac{\omega \times e}{n} \tag{16}$$

The above equation shows that the LHS which is marginal cost of testing (derivative of the cost function) is equal to the average cost of testing.

3.2 Cost of Additional Testing

Since the determinants of the cost of testing are the wages (ω) that need to be paid for the CT engineers hired $(e = {}^{E}/_{L})$. Cost of testing can be simply expressed as:

$$c = e \times \omega$$

From equation 16, it is known that the marginal cost of an additional round of testing is equal to the average cost of additional testing. Therefore, the marginal cost can be written as:

$$\underbrace{\Delta c}_{change\ in\ cost} = \underbrace{\frac{e \times \omega}{n}}_{\substack{Cost\ of\ testing \\ per\ round}} \times \underbrace{\Delta n}_{\substack{change\ in\ frequency \\ of\ testing\ in\ a\ year}} \tag{17}$$

Using the tax and spend equation (equation 10) and substituting in the above marginal cost equation, the expression is written as:

$$\boxed{\Delta c = \frac{s\tau}{(g + \delta + s\tau)} \times \frac{\Delta n}{n} \times y} \tag{18}$$

This formula estimates the change in cost resulting from change in the frequency of testing. The cost impact is scaled by:

- The proportion of output per worker allocated to testing.
- The proportional change in the amount of testing in a year.
- The output per worker.

3.3 Benefits of Additional Testing

For the benefits, we consider the production function of testing discussed in the background section with the effective total factor productivity. A key component considered here is that the output is proportional to the probability of success which corresponds to functioning capital. Recall that $y = (1-\theta)vAk^{\beta}$; y is the output per worker, θ is the regret rate (good output thrown away as defected), v is the reliability of production (functioning capital), A is theoretical TFP and Effective TFP is denoted by B where $B = (1-\theta)vA$ and $A > B$ and k is the capital intensity.

For simplicity, let's take the logarithm of the Cobb Douglas production function and then derivate it to compute the marginal benefits.

$$\log(y) = log\left[(1-\theta)v\mathrm{Ak}^{\beta}\right]$$

Using the logarithmic identity:

$$\log(y) = log(1-\theta) + \log(v) + \log(A) + \log(\mathrm{k}^{\beta})$$

Taking the derivative w.r.t time:

$$\frac{\Delta y}{y} = \frac{\Delta(1-\theta)}{1-\theta} + \frac{\Delta v}{v} + \frac{\Delta A}{A} + \beta\frac{\Delta k}{k}$$

Since *A* and k^{β} are fixed in the short run:

$$\frac{\Delta y}{y} = \frac{\Delta(1-\theta)}{1-\theta} + \frac{\Delta v}{v}$$

The marginal benefits equation can be written as:

$$\frac{\Delta y}{y} = \frac{\Delta v}{v} - \frac{\Delta\theta}{1-\theta} \qquad (19)$$

Substituting the expression for $\Delta\theta$ from equation 6 in the above expression:

$$\frac{\Delta y}{y} = \frac{\Delta v}{v} - \frac{\mathrm{p}\times(1-\theta)\times\Delta n}{1-\theta}$$

After rearranging, the marginal benefits equation be rewritten as:

$$\frac{\Delta y}{y} = \frac{\Delta v}{v} - (\mathrm{p} \times \Delta n) \tag{20}$$

Hence, the proportional change in output per worker (labour productivity) is impacted by the proportional change in the reliability of production and the probability of losing output in each additional round of testing to type I errors.

From equation 15, we have a developed expression for proportionate change in reliability of production $\frac{\Delta v^*}{v^*}$. Substituting that in the above equation, gives:

$$\Delta y^* = y^* \frac{\Delta n}{n} \left\{ \frac{qn(1-\phi)(1-v^*)}{(\delta + g + \phi + \varepsilon)v^*} - \mathrm{pn} \right\} \tag{21}$$

From equation 12, we know that $v^* = \frac{\delta+g+\phi}{\delta+g+\phi+\varepsilon}$ and so the marginal benefits equation can be rewritten as:

$$\boxed{\frac{\Delta y^*}{y^*} = \frac{\Delta n}{n} \left\{ \frac{qn(1-\phi)(1-v^*)}{\delta + g + \phi} - \mathrm{pn} \right\}} \tag{22}$$

The equation shows the net marginal benefit of an additional round of testing. It has two parts: one is gain from detecting bad output and the other is loss from scrapping good output. If you imagine servicing your car more frequently, the servicing helps because it removes emerging faults, but it also hurts because mechanics may sometimes replace parts unnecessarily and it is costly, so the equation effectively says that a 10% increase in servicing improves performance by (benefits minus losses) multiplied by that same 10%.

3.4 Optimal Level of Testing

Optimal level of testing is defined by the point where marginal benefits (Δy) equal the marginal costs (Δc). Let's bring both the expressions together.

$$y^* \frac{\Delta n}{n}\left\{\frac{qn(1-\phi)(1-v^*)}{\delta+g+\phi} - \mathrm{pn}\right\} = \frac{s\tau}{(g+\delta+s\tau)} \times \frac{\Delta n}{n} \times y^*$$

$$\frac{qn(1-\phi)(1-v^*)}{\delta+g+\phi} - \mathrm{pn} = \frac{s\tau}{(g+\delta+s\tau)}$$

$$\boxed{\frac{qn(1-\phi)(1-v^*)}{\delta+g+\phi} - \mathrm{pn} = m^*} \qquad (23)$$

Writing the expression as:

$$\frac{qn(1-\phi)(1-v^*)}{\delta+g+\phi} = \mathrm{pn} + m^*$$

It can be seen that the left-hand side represents the marginal benefit of extra testing which is the gain from detecting more bad output and resetting the part of capital that is malfunctioning. The right-hand side has two components:

- Additional scraping of output due to false positives – a result of more testing
- the cost of hiring CT engineers

The gain must offset these two costs for an additional round to be beneficial.

4 Parameters, Variables & Derived Quantities

4.1 Decomposition of Output

From the setup of the framework, it can be concluded that output has three categories:

- ***Success rate*** (Λ) – The success rate is the good output that is consumed by users and doesn't incur any error in the testing process.

$$\Lambda = (1 - \theta)v \tag{24}$$

- ***Rebate rate*** (Ω) - Defective output mistakenly enters the supply chain and is returned by users.

$$\Omega = (1 - \phi)(1 - v) \tag{25}$$

- ***Scrap rate*** (Γ) – Defective output correctly scrapped.

$$\Gamma = \phi(1 - v) + v\theta \tag{26}$$

The three types of output equate to 100% i.e. $\Lambda + \Omega + \Gamma = 1$

From the success and rebate rate equations, we can derive an expression that eliminates v. This will then be used in section 5 to derive the values of θ and ϕ.

$$\Lambda = (1 - \theta)v$$

$$\Omega = (1 - \phi)(1 - v)$$

$$\frac{1 - \theta}{1 - \phi} = \frac{\Lambda}{1 - \phi - \Omega}$$

$$\boxed{1 - \theta = \frac{\Lambda(1 - \phi)}{1 - \phi - \Omega}} \tag{27}$$

4.2 Table of "Knowns" and "Unknowns"

The model features a number of components some of which are reasonably estimated others are relatively challenging to estimate. King and Nayak (2025) classify them into three categories and covers this in great detail. The main categories of the components are basic parameter, variable or derived quantity. The table below summarises the main components that feature in this model and highlights an estimate where possible[10]. It further points out the components that need to be estimated using the golden rule of optimal testing.

For estimating the unknowns, it is assumed that the actual economy - the one that we currently inhabit - has settled into its equilibrium and then use the established equilibrium relationships to find the values of the unknown. The model here consists of a large set of interdependent quantities requiring estimation. To minimize the complexity of the model, algebraic simplification is used whereby the system is reformulated such that only two key quantities remain unidentified. These two quantities can be estimated directly from the system of simultaneous equations, and their values subsequently allow for the derivation of all the other unknown components. The estimates shown in the table below are based on statistics reported between 2015 and 2019 with 2017 as the midyear. Details of these calculations can be found in King and Nayak (2025).

Quantity	Symbol	Estimate	Notes
Basic Parameters			
wage rate	w	£33,000	
rental rate	r	11.20%	
share to labour	α	68.90% of GVA	
share to capital	β	31.10% of GVA	
Effective Total Factor Productivity (TFP)	B	10.47	$B = \frac{y}{k^{\beta}}$
workforce	L	25.10m	Number of productive workers
Savings rate	s	17.93%	$s = \frac{(\delta + g)}{(1 - m)} \frac{k}{y}$
Depreciation rate	δ	4.80%	
growth rate of the workforce	g	1.50%	

[10] The calculator covers each of the component and the process of calculating, in detail.

gross investment rate	$\delta + g$	6.30%	
statistical power of the test	q	99.90%	
probability of false positives	p	?	
pace of testing	a	?	
flip rate	ε	?	
wage of CT engineers	ω	£32,400	
Variables			
capital intensity	k	£133,944	
reliability of production	v	?	
Frequency of testing	n	?	
Derived quantities			
CT engineers (as % of workforce)	$e = E/L$	2.79%	
Portion size	ρ	£4.80m per engineer	
'Reliability tax'	τ	0.67%	
spend on conformance testing	m	1.88%	$m = \tau(k/y)$
detection rate	ϕ	?	
regret rate	θ	?	
power and frequency	qn	?	
Probability of type I and frequency	pn	?	
Elasticity of reliability of production w.r.t frequency	$\mathcal{E}(v, n)$	?	
scrap rate	Γ	3.70%	
rebate rate	Ω	1.50%	
success rate	Λ	94.80%	
yearly output (GVA)	Y	£1,203.80bn	
output per worker (labour productivity)	y	£47.96	
net output per worker	$y - \omega e$	?	

Table 1 Types of components in the model

5 Solving Simultaneous Equations and Some Comparative Statics

5.1 Solving Simultaneous Equations

The aim is to be able to solve the expression for n but because this is an implicit equation where ϕ depends on n and v which depends on ϕ, the route to take is to develop the expression further and eliminate all variables except for ϕ, θ and constants.

Since qn and pn are symmetric and $\phi = 1 - \exp(-qn)$, therefore, $\theta = 1 - \exp(-pn)$

Solving for qn by taking log on both sides, gives:

$$qn = \ln\left(\frac{1}{1-\phi}\right), pn = \ln\left(\frac{1}{1-\theta}\right) \tag{28}$$

Putting these expressions for qn and pn, and applying the notations discussed in section 5 equation 23:

$$\boxed{\left\{\frac{\Omega}{\delta + g + \phi}\right\} \ln\left(\frac{1}{1-\phi}\right) - \ln\left(\frac{1}{1-\theta}\right) = m^*} \tag{29}$$

Equation 27 and 29 can now be solved simultaneously for θ and ϕ.

5.2 Numerical Solution

The following components can be estimated once the values are obtained for the detection and regret rate.

Components	Estimate	Notes
detection rate	68.65%	Solving the two equations simultaneously
regret rate	0.44%	Solving the two equations simultaneously
reliability of production	95.21%	$v^* = 1 - \frac{\Omega}{(1-\phi)}$
Frontier TFP	11.03	$\mathrm{A} = \frac{y}{(1-\theta)vk^{\beta}}$
frequency of testing	1.16	$n = \frac{1}{\mathrm{q}} \ln\left(\frac{1}{1-\phi}\right)$
power and frequency	1.16	$qn = \ln\left(\frac{1}{1-\phi}\right)$
probability of false positives	0.38%	$n = \frac{1}{p} \ln(\frac{1}{1-\theta})$
Probability of type I and frequency	0.437%	$pn = \ln(\frac{1}{1-\theta})$
Elasticity of reliability of production w.r.t frequency	2.322%	$\frac{n}{v^*}\frac{dv^*}{dn}$
pace of testing	£5.57m	$a = n \times \rho$
flip rate	3.77%	$\varepsilon = \frac{\phi \times (1-v^*) + (\delta + g) \times (1-v^*)}{v^*}$
Relative standard deviation	17.35%	$\frac{1}{z(1-p) - z(1-q)}$

Table 2: Calculations

Increasing the frequency of inspections will improve the reliability of production, v. The calculations show that doubling the frequency of inspections would increase the reliability of production by about **2.4%**. That is, the elasticity of v with respect to n. Intuitively, the "golden rule" for the optimal amount of testing is based on equating the cost of doubling the frequency of testing (again) with the benefit of doubling the frequency of testing (the elasticity of v with respect to n minus further losses from false positives). Society should keep doubling the amount of testing until it reaches a threshold where if testing were to double one more time, then the increased cost matches the increased benefit. Up to this threshold level of testing, an increase in testing will always add more in benefits than it takes in costs. The Golden rule optimality condition can be written as follows:

$$\underbrace{\mathbf{1.88\%}}_{\substack{\text{Spend on testing}\\ \text{as a}\\ \text{percent of output}}} = \underbrace{\mathbf{2.322\%}}_{\substack{\text{Elasticity of reliability}\\ \text{w.r.t}\\ \text{frequency of testing}}} - \underbrace{\mathbf{0.437\%}}_{\substack{\text{Losses from}\\ \text{false positives}\\ \text{(type I errors)}}}$$

The “golden rule” is based on equating the marginal cost and marginal benefit of the extra information from another round of testing. Doubling the frequency doubles the cost (2% of output again). Doubling the frequency increases the reliability of production processes but adds to the loss from false positives in the testing process.

The result for frequency of testing implies that instruments should be tested every 10 to 11 months to get the optimal performance from the instruments. It can also be explained using the example of servicing cars. Most cars are designed to be serviced roughly every 12 months or 12,000 miles, a recommendation commonly cited by sources such as RAC. This interval is not arbitrary; manufacturers deliberately set it to strike a balance between cost and reliability. Servicing too frequently can lead to unnecessary expense for the owner, while delaying maintenance increases the risk of mechanical failure and costly breakdowns. As a result, the recommended schedule reflects a careful optimisation of factors including normal wear and tear, the probability of component failure over time, and the overall trade-off between routine servicing costs and potential repair expenses.

Without the existence of a reliable measurement infrastructure i.e. having a traceable calibration chain, the effectiveness of additional testing diminishes. This is shown in Figure 4 explained in section 5.3, where an increase in the probability of type I error (p) will shift the marginal benefits curve downwards and thereby, shifting the optimal levels down.

5.3 Comparative Statics using the Calculator

Having done the calculations and finding all the components including the basic parameters. We can now go the other way and perturb one of the basic parameters shift while keeping the others the same to see the impact on the other components; also known as comparative statics. Let that basic parameter be the probability of type I error, which shifts from 0.375% to 0.44%. This refers to the counterfactual where NMS labs no longer exist and so, access to precise calibrations becomes limited. The table below summarises the components that change in this new equilibrium and compares them to the baseline. [11]

Components	Baseline scenario which finds $p = 0.376$	New scenario with $p = 0.441$	Percentage change
Capital intensity (k) (unit: thousands)	£133.94	£133.80	-0.1%
Output per worker (y) (unit: billions)	£47.96	£47.89	-0.1%
Output (Y) (unit: billions)	£1,203.80	£1,202.16	-0.1%
Cost of Testing (mY) (unit: billions)	£22.69	£22.32	-1.6%
Net Output (Y-mY) (unit: billions)	£1,181.11	£1,179.84	-0.1%
Effective TFP	11.031	11.031	0.0%
Observed TFP	10.457	10.446	-0.1%
Detection rate	68.652%	68.098%	-0.8%
Regret rate	0.436%	0.503%	15.4%
Success rate	94.800%	94.702%	-0.1%
Rebate rate	1.500%	1.537%	2.5%

[11] Comparative statics compares different equilibria after a parameter shift. Sensitivity analysis studies how sensitive the current equilibrium is to small changes in parameters.
Here, the variation in p is taken as exogenous, but it is grounded in the Gravity Model for calibration uptake (see King and Nayak, 2025), which is not developed further in this report but will be unpacked in detail in a subsequent study.

Baseline likelihood of false positives	p	0.376%
Relative Standard Deviation (RSD)	σ	17.352%
% change in accuracy	$\Delta\sigma/\sigma$	1%
Change in likelihood of false positives	Δp	0.065%
New likelihood of false positives	$p + \Delta p$	0.441%

Probability of type I error	0.376%	0.441%	17.2%
Flip rate	3.767%	3.767%	0.0%
Reliability of production	95.215%	95.181%	0.0%
frequency of testing (yearly)	1.161	1.144	-1.5%
Elasticity of reliability of production w.r.t frequency of testing	2.322%	2.361%	1.7%
Relative Standard Deviation	17.352%	17.526%	1.0%

Table 3: Perturbation of the probability of type I errors

In the absence of NMS labs, i.e. a higher type I error rate, cause the net output to decline by 0.1%. This is because components like the regret rate increases fairly significantly which means more output is being scrapped erroneously. Detection rate can be seen to slightly decrease implying less output is being detected as defected and so the part of capital is less likely to be reset.

5.4 Graphical Explanation of Optimal Testing

The figure below provides a diagrammatical explanation of the golden rule for optimal testing.

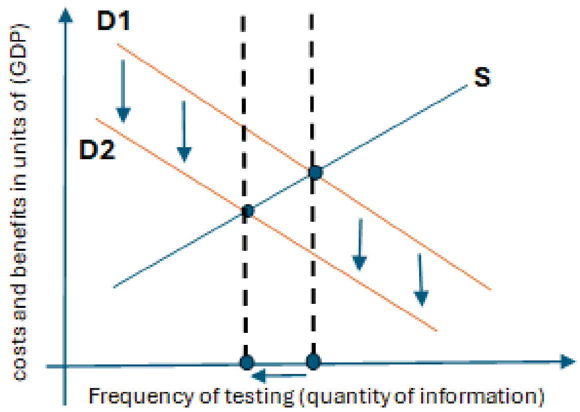


Figure 5 Optimal Level of Testing

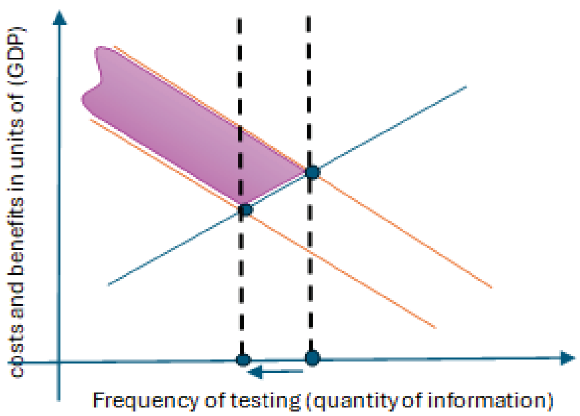


Figure 6 Effect of decrease in testing on Economic surplus

The downward sloping orange lines (D1 and D2) are two different demand curves for testing derived from the marginal benefit of an additional round of tests. The downwards shift from D1 to D2 represents an increase in the likelihood of false positive (type 1 errors) in the testing process which makes testing less effective and raises the scrap rate. The upward sloping blue line (S) is the supply curve for testing derived from the marginal cost of additional rounds of testing. The point where the demand curve and supply curve intersect defines the point of equilibrium. The sketch shows that if the demand curve moves down (due to an increase in the likelihood of false positives), then the equilibrium level of testing will decline, as represented by the horizonal dashed line moving to the left.

Figure 5 shows what happens to the surplus (net benefit) when following a downward shift in the demand curve (due to an increase in the likelihood of false positives). The area under the demand curve (orange line) but above the supply curve (blue line) gives economic surplus created by using some resources to perform tests and reset malfunctioning production processes. Hence, the purple area in the sketch above represents the decrease in the surplus following a downward shift in the demand curve and a decrease in the equilibrium level of testing.

6 Conclusion

This report discusses the macroeconomic model presented in Mike and Nayak (2025) for identifying the optimal level of conformance testing in an economy, where production processes are imperfect and capital can sometimes malfunction. By adapting the Golden Rule Savings rate logic from the Solow–Swan model, the level of testing at which the marginal benefits exactly equal the marginal costs of additional testing, is derived. The golden rule maximises consumption per worker which is also where capital intensity and output is maximised. Several key results emerge from the model analysis:

- Testing improves productivity by increasing the share of functioning capital and reducing the flow of defective output into the economy.
- The economy reaches its highest sustainable performance at a testing frequency where the next round of testing adds no further net value.
- Excessive testing becomes counterproductive, raising Type I errors and increases the 'reliability tax' that depresses both net output and steady-state capital.
- Even small changes in the probability of measurement error have system-wide consequences, affecting reliability, capital intensity, labour productivity, and output.

By reducing the complex model into a system of simultaneous equations, the calculator finds all the unknown components including the basic parameters. The model finds the optimal frequency of testing to be 1.16 (testing output every 10-11 months to ensure the machinery/ capital is functioning). More importantly it finds that testing frequency improves capital reliability, i.e. if testing frequency is doubled, the reliability of production increases by roughly 2.4%.

There is evidence from the NMS Private Businesses Survey that some customers of the NMS laboratories seek products and services abroad, suggesting that they cannot always obtain the required services within the UK. This indicates potential scope for improvement in domestic provision. However, the model used in this report abstracts from these frictions and assumes that the UK operates at the technological frontier with fully effective measurement services. Hence, there is scope to refine the model.

Several avenues for further development naturally follow:

- Use the calculator and conduct sensitivity analysis to see how small variations in basic parameters can alter other components. There may be scope to explore a tipping point analysis. [12]
- Update the calculator using estimates based on data for 2021-2023. This work is currently underway, and an accompanying report presenting the revised estimates and a more detailed theoretical explanation of the components is expected to be published in September 2026.
- With the optimal testing model complete, the next phase is to embed it within the wider measurement infrastructure framework i.e. integrating it with the **Gravity Model for Calibration Uptake**. This would quantify the implications of the absence of a domestic National Metrology Institute (NMI) implying reduced access to precise calibrations.
- Develop the **Calibration and Fanout Model** which links the traceability chain into our wider measurement infrastructure framework.

As a next step, the framework could be adapted to a manufacturing-specific version of the model. This would involve analysing a scenario in which only the manufacturing sector undertakes conformance testing, allowing for a more detailed assessment of testing activity within production-intensive industries. The analysis could also be extended to incorporate the sectors that provide outsourced testing services to manufacturers, thereby capturing the wider supply chain of testing, inspection, and certification activities. Such an extension would provide a more complex picture of the economy i.e. a multisector general equilibrium.

[12] A tipping point analysis is a statistical sensitivity test used to evaluate the robustness of a study or model's conclusions. It determines how much the missing data or underlying assumptions would need to change in order to overturn the model's primary findings.

7 References

1. Arrow, K.J. (1962) 'Economic welfare and the allocation of resources for invention', in National Bureau of Economic Research The Rate and Direction of Inventive Activity: Economic and Social Factors. Princeton: Princeton University Press, pp. 609-626
2. Dixit, A.K. and Pindyck, R.S. (1994) Investment under Uncertainty. Princeton: Princeton University Press.
3. Fennelly, C. (2021) Quantifying measurement activity in the UK. NPL Report IEA 7. Teddington: National Physical Laboratory. Available at: https://eprintspublications.npl.co.uk/9064/1/IEA7.pdf
4. Katanguru, D.R., Qureshi, Z. and King, M. (2025) A survey of UK-based businesses using laboratories funded through the National Measurement System (2023). NPL Report IEA 28. National Physical Laboratory. Available at: http://eprintspublications.npl.co.uk/10151.
5. King, M. & Nayak, S. (2025) A macroeconomic model with imperfect production processes and conformance testing. NPL Report IEA 27. Teddington: National Physical Laboratory. Available at: https://eprintspublications.npl.co.uk/10127/
6. King, M. & Nayak, S. (2023) An economic model for the value attributable to high-quality calibrations by reducing mistakes in conformance testing. NPL Report IEA 19. Teddington: National Physical Laboratory. Available at: https://eprintspublications.npl.co.uk/9816/
7. MacMinn, R.D. and Holtmann, A.G. (1983) 'Technological uncertainty and the theory of the firm', Southern Economic Journal, 50(1), pp. 120-136.
8. PA Consulting (1999) Review of the Rationale for and Economic Benefit of the UK National Measurement System. Available at: https://webarchive.nationalarchives.gov.uk/ukgwa/20070603154510/http:/www.dti.gov.uk/nms/policy/pareview.pdf
9. Qureshi, Z. & King, M. (2025) A framework for assessing the economic significance of NPL's spillovers. NPL Report IEA 25. National Physical Laboratory. Available at: https://doi.org/10.47120/npl.IEA25
10. Sims, C.A., 2003. Implications of rational inattention. Journal of monetary Economics, 50(3), pp.665-690

11. Swann, G.M.P. (2000) The Economics of Standardization. Report for the UK Department of Trade and Industry. Manchester Business School. Available at: https://library.iso.org/contents/data/238-the-economics-of-standardiza.html
12. Swann, G.M.P. (2009) The Economics of Measurement and Standards. National Measurement System Report, Department for Business, Innovation and Skills.
13. Tassey, G. (1982) 'The Role of Government in Supporting Measurement Standards for High-Technology Industries,' Research Policy, 11, pp. 311–320.
14. Tassey, G. (1982) 'Infratechnologies and the role of government,' Technological Forecasting and Social Change, 21(2).
15. Ye, F., Paulson, N. and Khanna, M. (2024) 'Strategic innovation and technology adoption under technological uncertainty', Journal of Economic Dynamics and Control, 165, p. 104879. Available at: https://www.sciencedirect.com/science/article/abs/pii/S016518892400071X

Annex A: Guide for the Calculator

Link to the Calculator

The calculator is designed to show a mechanical way of how the Optimal Testing Model (OTM) operates. The calculator constructs two scenarios. The first is the **actual scenario – the true equilibrium of that world**. This scenario corresponds to when traceable calibrations are provided by the NMS labs (domestic NMI) and so the economy has access to precise calibrations. The second is the **counterfactual, which represents the absence of these precise measurements provided by the NMS labs.**

To construct the actual equilibrium, the calculator begins with what the OTM calls the golden rule. The golden rule captures the frequency of testing where the marginal benefit from an additional round of testing equates the marginal cost from the additional round. The complex system of equations reduces to two equations with two unknowns (detection and the regret rate) which are solved simultaneously. These equations (27) and (29) of the Optimal Testing Report link the detection and the regret rate as seen below:

$$\boxed{1-\theta=\frac{\Lambda(1-\phi)}{1-\phi-\Omega}}$$

$$\boxed{\left\{\frac{\Omega}{\delta+g+\phi}\right\}\ln\left(\frac{1}{1-\phi}\right)-\ln\left(\frac{1}{1-\theta}\right)=m^{*}}$$

Using the golden rule and the 'goal seek' function of excel the values for the detection and regret rate are computed. The basic parameters of the model are inferred from these values and hence, the equilibrium components of the model are computed. This refers to the 'actual' scenario for the equilibrium.

The calculator then goes the other way and perturbs the probability of type I error (p) found in the equilibrium, to construct a counterfactual where precise calibrations are no longer accessible. This part allows us to see the impact on the system as p increases. It is important to note that the basic parameters are kept fixed and not changed here.

Explanation of the tabs of the Calculator

- The first three tabs provide a top-level description and overview of the components of the model. 'Components detailed' sets up the calculator highlighting the known and unknown components.
- The baseline equilibrium has three designated tabs, one for the basic / actual scenario, one for the calculation of the detection and regret rate using the 'goal seek' function and last for the counterfactual scenario where the value of the probability of type I error (p) is perturbed.
- 'EQ1' tab shows the goal seek function which calculates detection and regret rate. Goal Seek finds the value of ϕ that makes $\mathrm{F}(\phi) = 0$. Once found, these values automatically populate the unknown basic parameters and other components of the model e.g. the frequency of testing, probability of type I error etc.
- 'Change in p – EQ1' refers to the counterfactual where the NMS labs no longer exist and cause the relative standard deviation to increase which ultimately increases the probability of type I error. This tab can be treated as a perturbation in the value of p (comparative statics) where the basic parameters found in the 'Calculator EQ1' are kept fixed. In this tab the function for the testing frequency (n) is defined and 'goal seek' function is used to find the value of n that makes the $F(n) = 0$. This value will update all the other components.
- By comparing the actual and counterfactual, the impact on the system can be derived.
- The 'Comparison Scenarios' tab compares all the actual and the counterfactual scenario comparing for instance, output per worker, capital intensity, net output per worker etc.

Annex B: Overview of Remaining Models within the Framework

Gravity Model

NPL provides calibrations traceable to national standards

Organisations coming to NPL for calibrations represent the first link of the traceability

To model the shift in the uptake due to distance is done by imagining if somehow NPL traceability chain

The conventional formula for a country's internal distance is $DD = D = 0.67\sqrt{A/_{\pi}}$ where A is the geographical area of a country. The geographical area covered by the UK is 244,376 km2, which implies that its internal distance is 185.8 km.

King & Renedo (2020) find that if the time required to travel between the UK and another country was somehow too double, then demand for NPL's services would drop by 48%. In other words, the coefficient for logged distanced in the regression function is **0.48**.

The effect of switching off NPL's services is assessed by considering the increase in the distance that UK-based customers would need to ship any instruments that they wanted precisely

The distance from London to Amsterdam is 357 km.

The consequent drop in usage by UK-based customers is estimated as follows:

$$-0.48 \times [ln(357) - ln(186)] \approx -31\%$$

If UK retains access of access to precise calibrations through VSL, then the demand for precise calibration would drop by 31%. This reduction in access would lead the RSD to increase by about **one percentage point,** causing the probability of type I error to increase **by 0.05%.**

Calibrations and Fanout Model – Work-in-progress

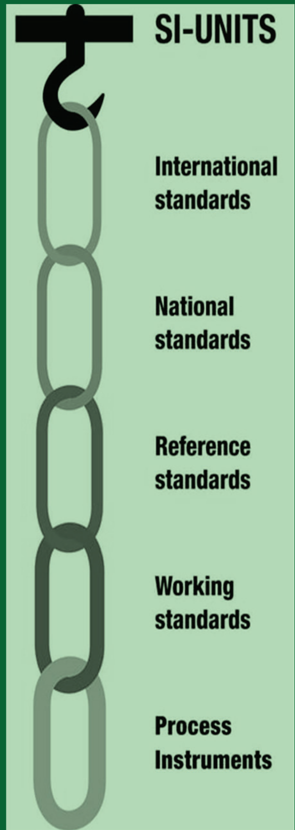
A model for understanding the distribution and transportation of information within

Traceability chain has 4 tiers

1. The SI Definition & National Metrology Institutes (NMIs)
2. Primary Reference Laboratories
3. Secondary Industrial Calibration Laboratories
4. The End User / Process

The model has 3 states for an instrument in each of the tiers

1. **State A (Out of Service):** Instruments that have been sent to a higher-tier calibration lab for recalibration.
2. **State B (In Service and Properly Calibrated):** Instruments that are actively used to perform calibrations and are properly calibrated.
3. **State C (In Service but Not Properly Calibrated):** Instruments still in use but no longer properly calibrated, producing unreliable calibrations.



At each tier, the entities make just normal profit so there isn't any incentive to move up or down the chain.

instruments in state B and C are different such that $r_B > r_C$ but remains the same across all the tiers. These capture the opportunity cost for the instruments had they been performing testing and analysis.

How is the calibration chain sustained in equilibrium?

How would a change in the turnaround times for NPL's customers feed through to changes in the availability and quality of traceable calibrations for end-users at the bottom of the chain?

What fee should NPL be charging level 1 labs for its services?

As you move down from the NMI, each level acts as a multiplier. The number of instruments expands rapidly, while the price points compress due to commoditization and increased competition.

The rental rate for instruments in state B and C are different such that $r_B > r_C$ but remains the same across all the tiers. These capture the opportunity cost for the instruments had they been performing testing and analysis.

The turnaround time is defined as the time it takes at each level to produce a calibrated instrument.

$$\boxed{Turnaround\ time\ =\ queuing\ time\ +\ laboratory\ time\ +\ transportation\ time}$$

‘Basic’ & ‘Precise’ Calibrations Model

A measurement error has two components

‘**Random**’ error and ‘**Systematic**’ error.

Random errors can be averaged out by repeating measurements. **Systematic errors (calibration related uncertainty)** can only be eliminated/ minimised by **precise calibrations**.

The model uses the concept of TAR for the systematic error that comes with a basic calibration.

‘TAR is the comparison between the accuracy of a tool and the reference standard used to calibrate it. Metrology labs aim for a minimum TAR of 4:1, meaning the standard should be four times more accurate than the tool’

Precise calibration can eliminate all the systematic error, but basic calibrations cannot fully eliminate the component of the error associated with calibrations. It can be as much as 25% according to the concept of TAR

σ is expressed in terms of the size of the deviation that occurs under the hypothesis that the parts are defective. σ is thought of as the Relative Standard Deviation (RSD) of the measurement process. The standard deviation depends on the expanded uncertainty of the measurement process, incorporating a mixture of random and systematic sources of error.

Without access high-quality calibrations, businesses default to using a TAR of 4:1, implying that ${}^{u}/_{\sigma} = 25\%$.

Meaning, without a network of calibration labs taking traceability from national standards, σ would increase by 3.1%, leading to more mistakes in conformance testing.

$$\frac{\Delta\sigma}{\sigma_0} \approx \frac{1}{2}\left(\frac{1}{4}\right)^2 \approx 3.1\%$$

Type I and Type II errors Model

The model finds an expression connecting changes in σ to changes in the likelihood of type-1 (p) and type-2 errors ($1-q$).

$$\frac{1}{\sigma} = \Phi^{-1}(1-p) - \Phi^{-1}(1-q)$$

Assumption: There cannot be a change in q even if σ increases. The effect will wash up in p.

Type I error is known as a 'false positive'. Rejecting output that correctly conforms.

Type II error is known as a 'false negative'. Accepting output that does not conform.

'The relative standard deviation (RSD), σ, represents potential variation in the measured values for a single sample as a proportion of the required tolerance.

	Functioning Capital (Good Output)	Malfunctioning Capital (Bad Output)
Pass	$\Pr(\text{pass} \mid \text{good}) = 1-p$	$\Pr(\text{pass} \mid \text{bad}) = 1-q$
Fail	$\Pr(\text{faill} \mid \text{good}) = p$	$\Pr(\text{fail} \mid \text{bad}) = q$

Optimal Testing in the Economy

Providing the measurement capability that underpins the UK's prosperity

To find out more about NPL:

npl.co.uk

To get in touch:

npl.co.uk/contact

020 8977 3222

Hampton Road
Teddington
Middlesex
TW11 0LW

Please consider the environment before printing this document